High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors

P. Packan, S. Akbar, M. Armstrong, D. Bergstrom, M. Brazier, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He*, R. Heussner, R. James, J. Jopling, C. Kenyon, S-H. Lee, M. Liu, S. Lodha, B. Mattis, A. Murthy, L. Neiberg, J. Neirynck, S. Pae*, C. Parker, L. Pipes, J. Sebastian, J. Seiple, B. Sell, A. Sharma, S. Sivakumar, B. Song, A. St. Amour, K. Tone, T. Troeger, C. Weber**, K. Zhang, Y. Luo, S. Natarajan

Logic Technology Development, * Quality and Reliability Engineering, ** TCAD

Intel Corporation

IEDM 2009

Outline

- Scaling Trends
- 32nm Technology Features
- Device Performance
- Variation and Vccmin
- Summary

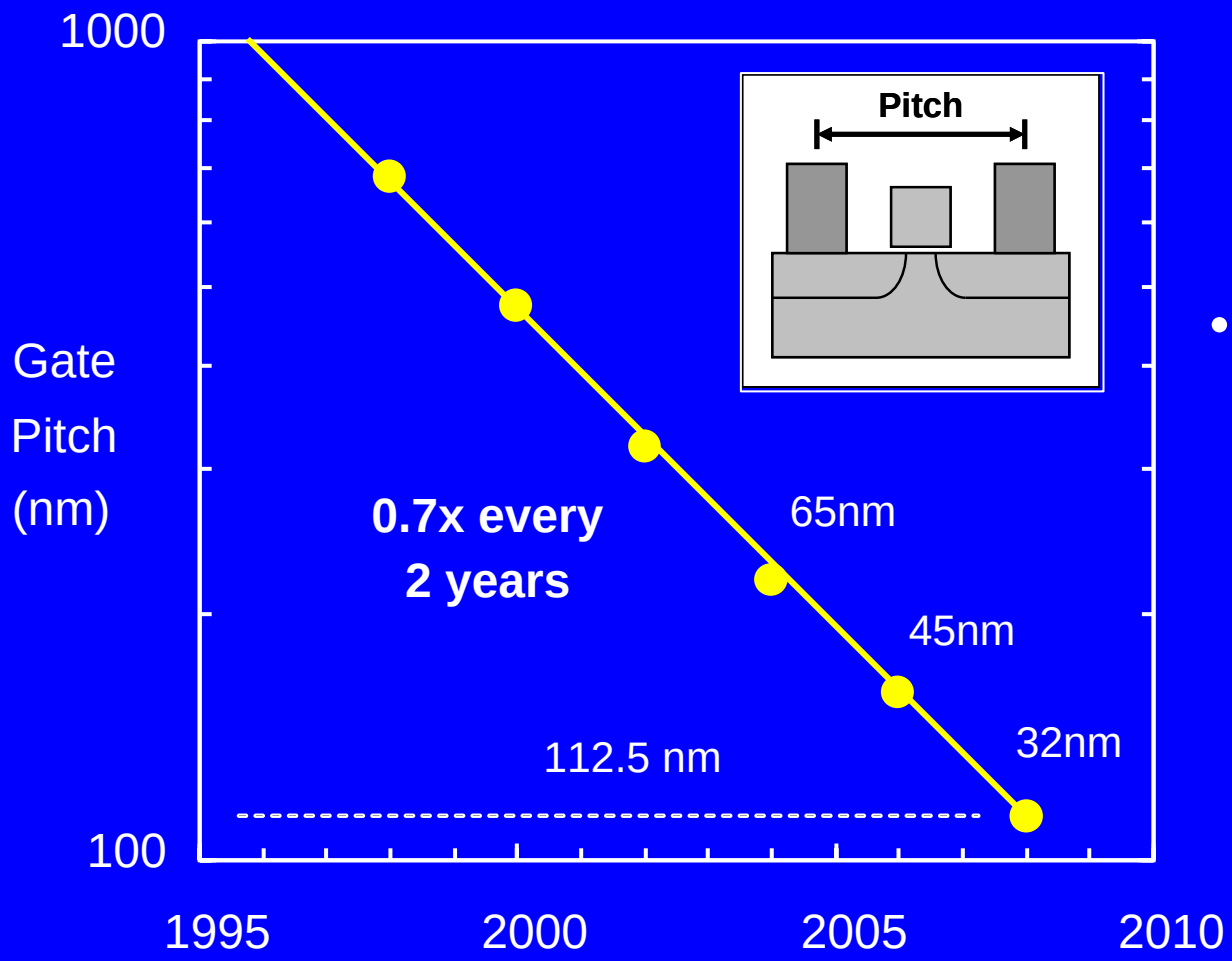
Key Messages

- 32nm technology continues historic scaling trends
  - Reduced pitch and increased performance
- Record NMOS and PMOS drive currents
  - NMOS Idsat of 1.62mA/um @100nA/um Ioff, 1.0V
  - PMOS Idsat of 1.37mA/um @100nA/um Ioff, 1.0V
- Highest reported SRAM array density for any 32nm or 28nm technology

Outline

- **Scaling Trends**
- 32nm Technology Features
- Device Performance
- Variation and Vccmin
- Summary

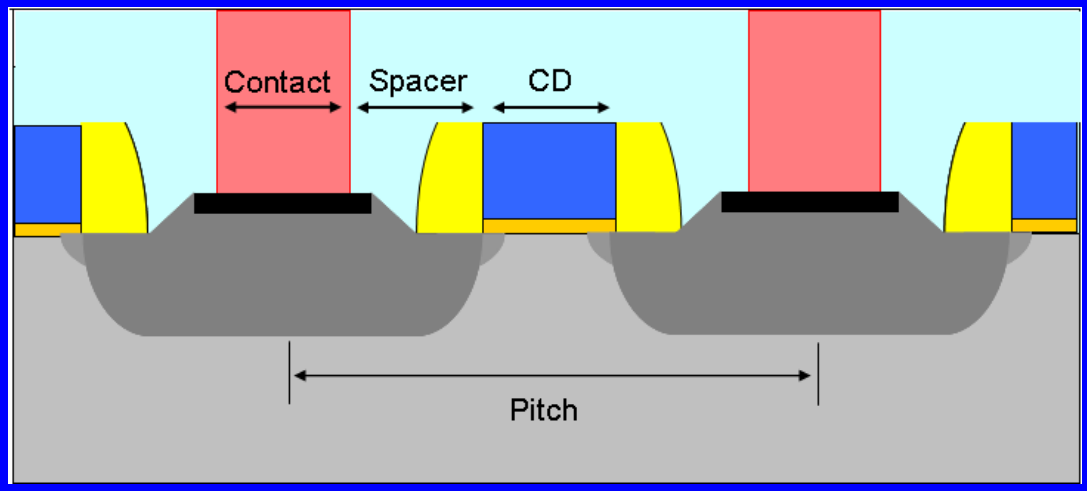
Pitch Scaling Trend

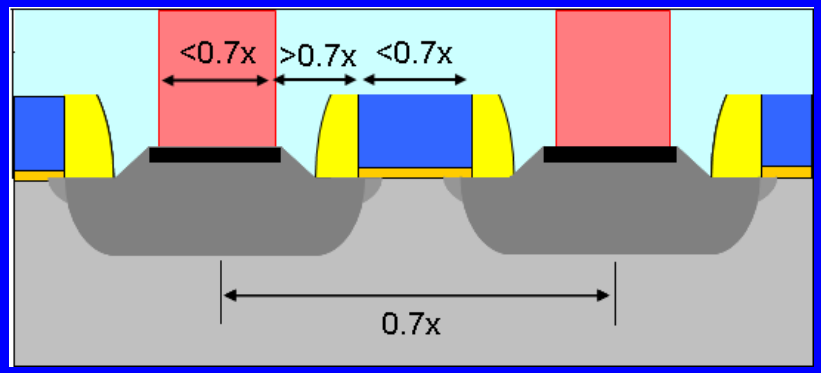


- Pitch scaling trend is continued to the 32nm technology node

Tightest reported gate pitch for any 32nm or 28nm node

Pitch Scaling Issues



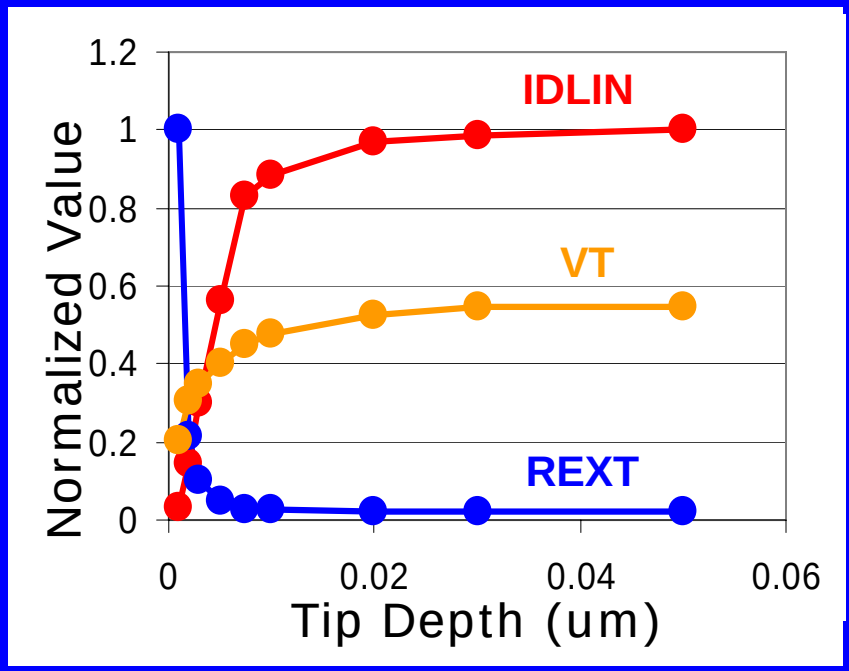


- Reduced contact area increases resistance
- Reduced area for strain enhancement
- Reduced gate length increases Vt
- Increased fringe capacitance

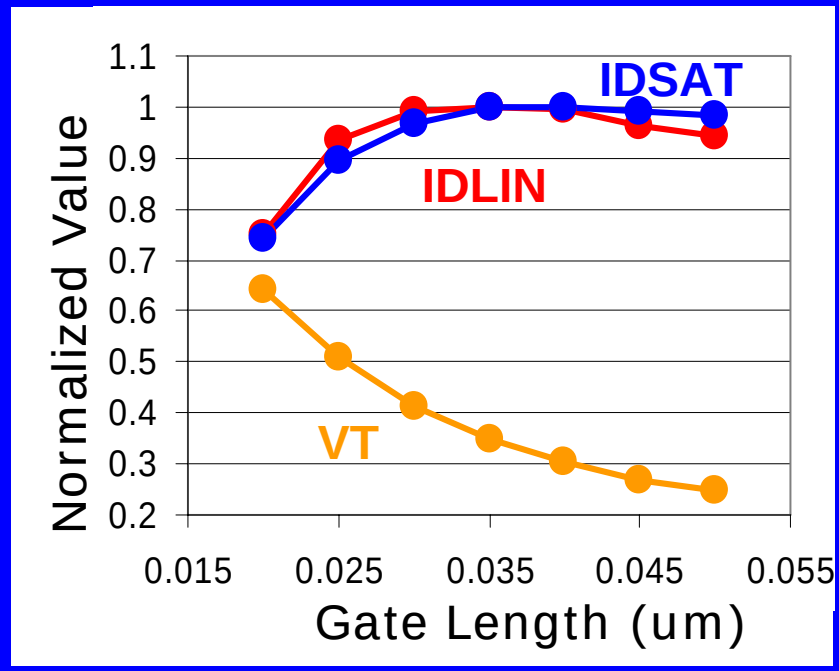
Novel techniques needed to scale pitch and improve performance

Traditional Scaling

$Id=\mu Cox/Le(Vg-Vt)^{\alpha}$



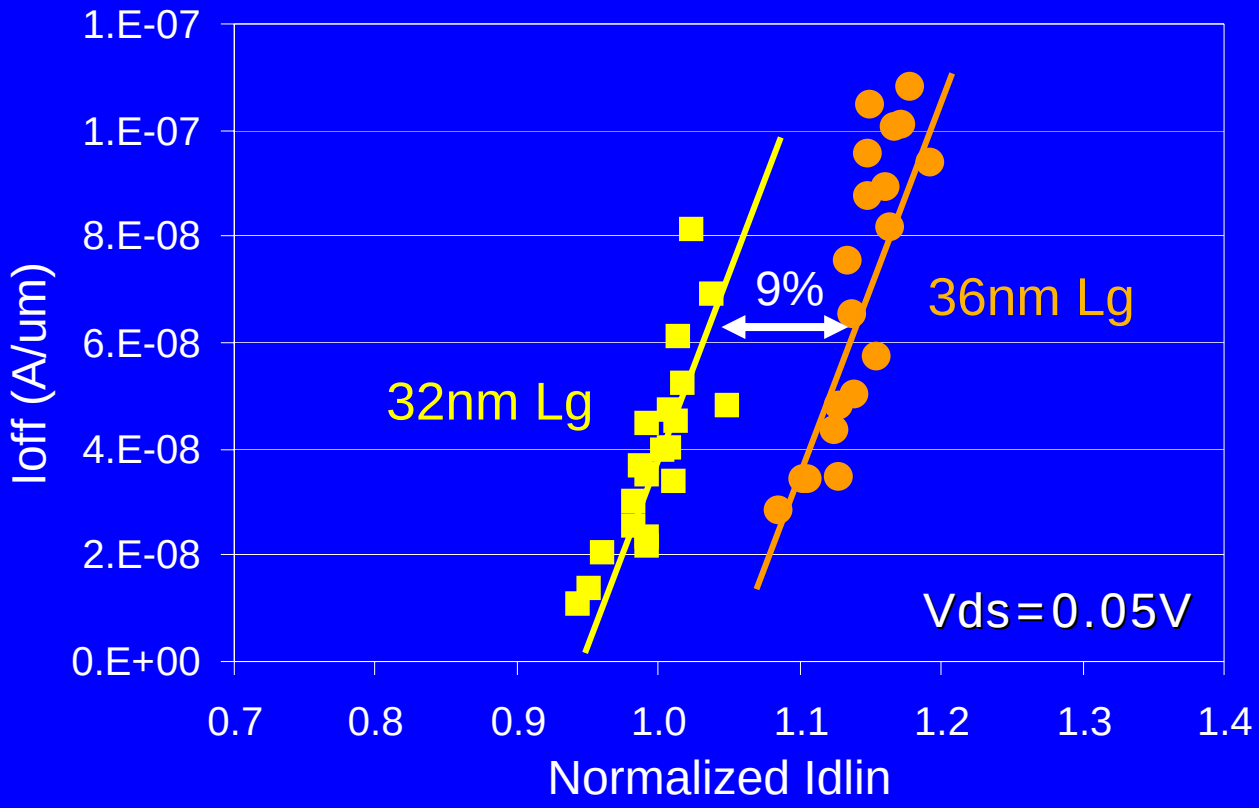
Junction Scaling



Vt Scaling

- Junction scaling is slowing due to resistance increases
- Gate length scaling through increased Vt degrades performance
- Traditional scaling is losing steam – new paradigm needed

Gate Length and Drive Current

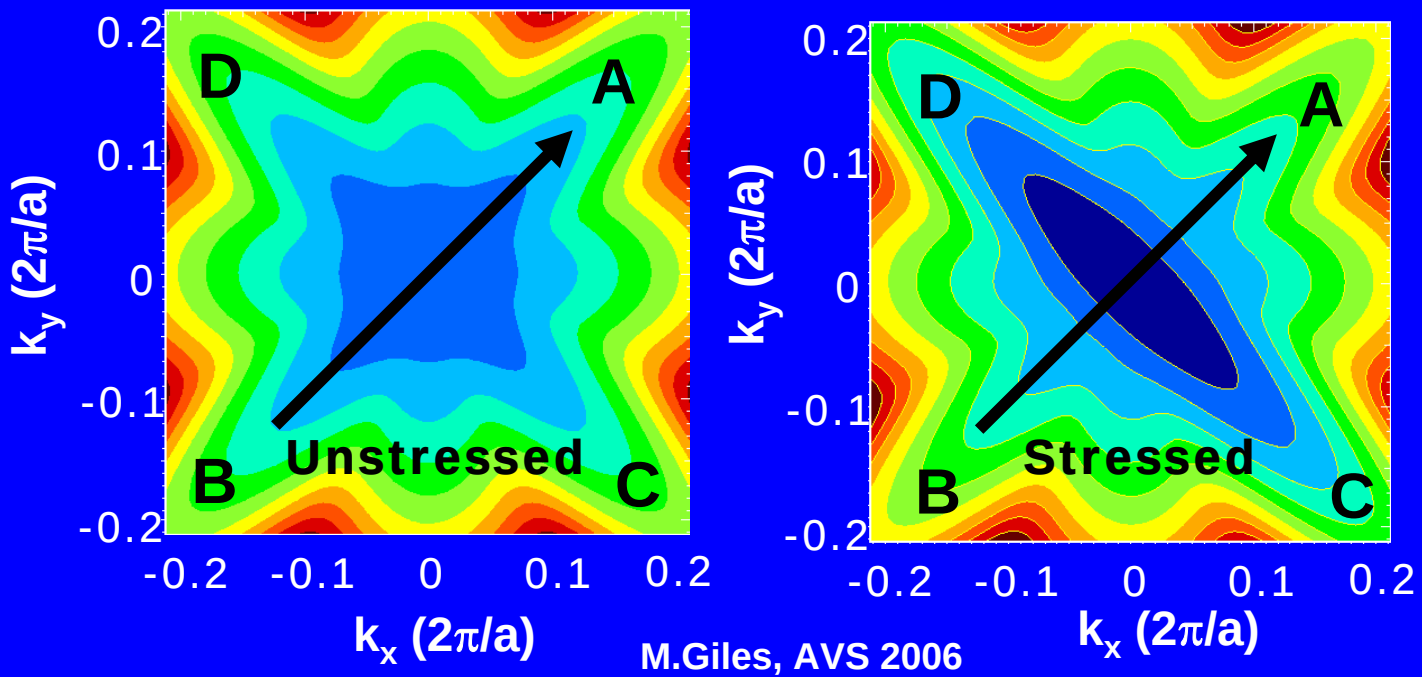


- Longer gate lengths can improve drive current for the same Ioff due to the lower Vt
- However, density and capacitance benefits at the shorter gate lengths are still preferred

Mobility Scaling

$Id=\mu Cox/Le(Vg-Vt)^{\alpha}$

- Increasing mobility increases device performance with minimal impact to leakage – effective scaling strategy
- Reducing scattering mechanisms, applying stress and surface orientation all affect mobility



M.Giles, AVS 2006

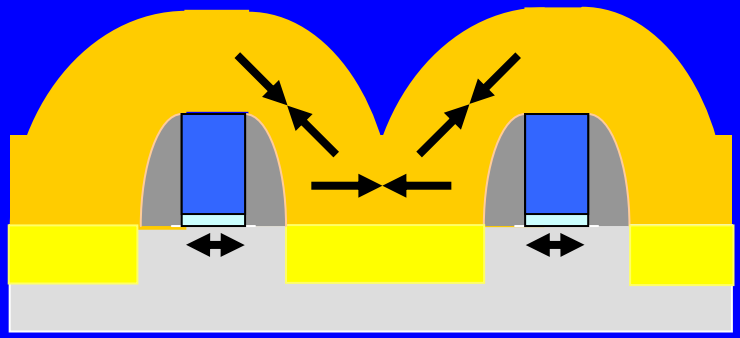
(001) surface

1GPa uniaxial stress [110]

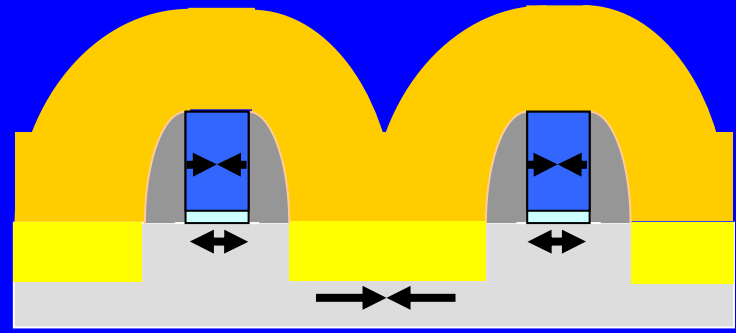
1 MV/cm vertical field

30meV energy contours

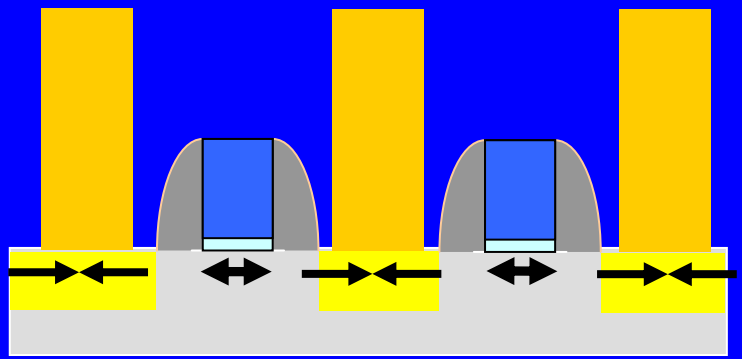
Stress Methods



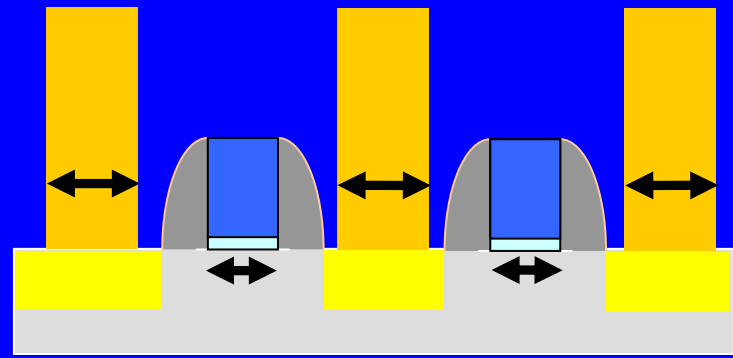
Capping Layers



Gate Induced



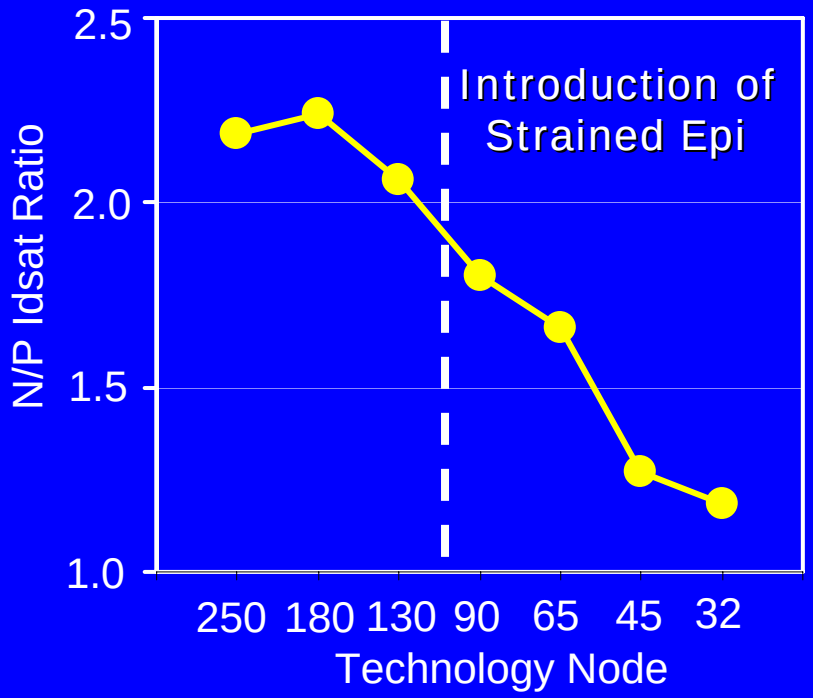
Epitaxial Layers

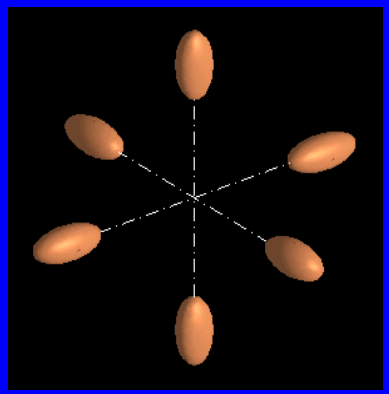


Contacts

C. Auth, VLSI 2008

NP Ratio Trend





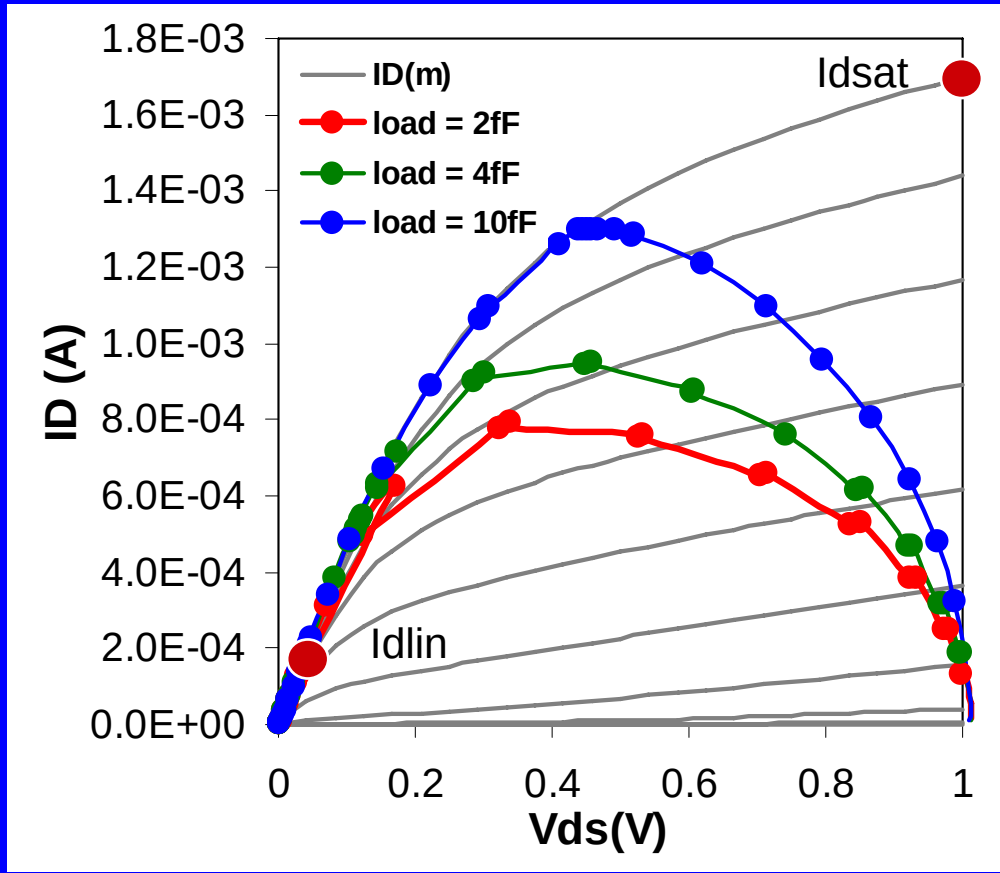
Electrons

Holes

- PMOS drive strength has been increasing more rapidly than NMOS primarily due to mobility improvements from stress
- Hole mobility shows a greater sensitivity to stress for (100)
  - Valence band degeneracy
  - Band energy splitting
  - Band warping
- Affects device sizing in circuits
  - Balances inverter circuits
  - Reduces PMOS area
- Improves performance and power

Idlin versus Idsat

- As an inverter switches, the transistor bias changes mapping out a current trajectory
- For typical capacitive loads, the majority of charging occurs at lower Vds values
- Based on these trajectories, Idsat may not be a good figure of merit for performance
- Idlin and Idsat are equally important for circuit performance



Outline

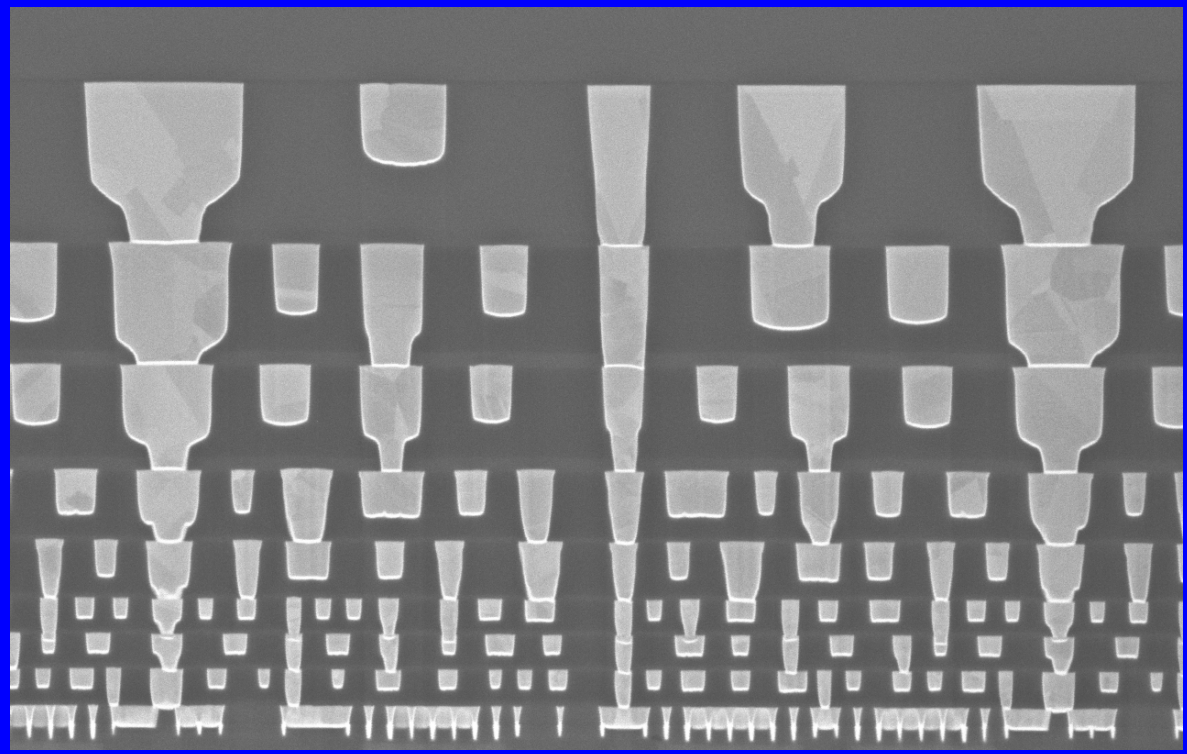
- Scaling Trends
- **32nm Technology Features**
- Device Performance
- Variation and Vccmin
- Summary

32nm Design Rules

Layer	Pitch (nm)	Thick (nm)	Aspect Ratio
Isolation	140	200	--
Contacted Gate	112.5	35	--
Metal 1	112.5	95	1.7
Metal 2	112.5	95	1.7
Metal 3	112.5	95	1.7
Metal 4	168.8	151	1.8
Metal 5	225.0	204	1.8
Metal 6	337.6	303	1.8
Metal 7	450.1	388	1.7
Metal 8	566.5	504	1.8
Metal 9	19.4um	8um	1.5

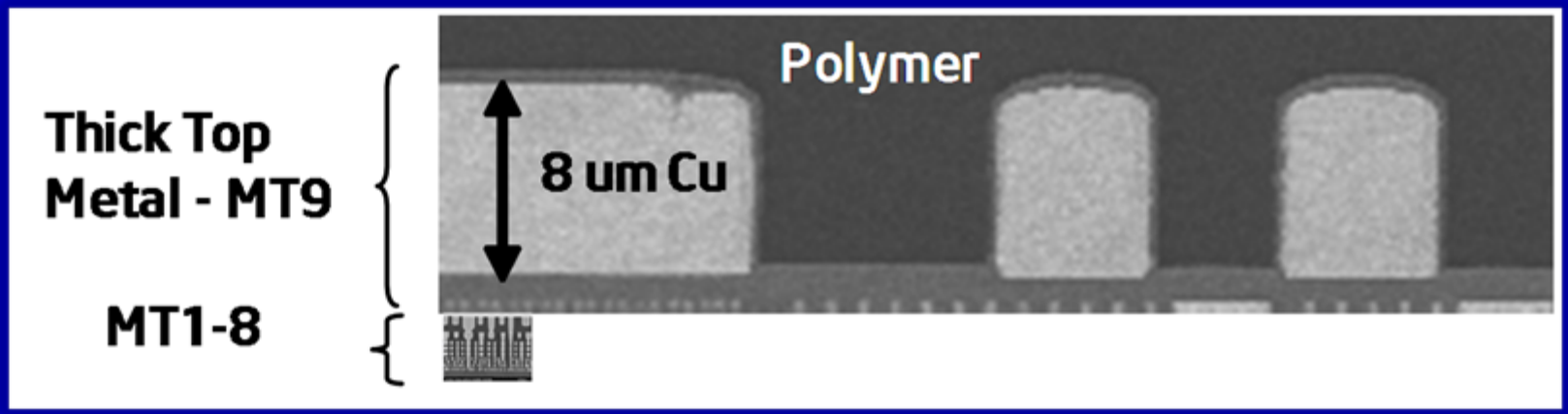
~0.7x scaling from 45nm technology

Interconnects



- 9 metal layers
- Metal 1-3 pitches match transistor pitch
- Graduated upper level pitches optimized for density and performance
- Extensive use of low-k ILD and SiCN materials

Thick Top Metal Layer

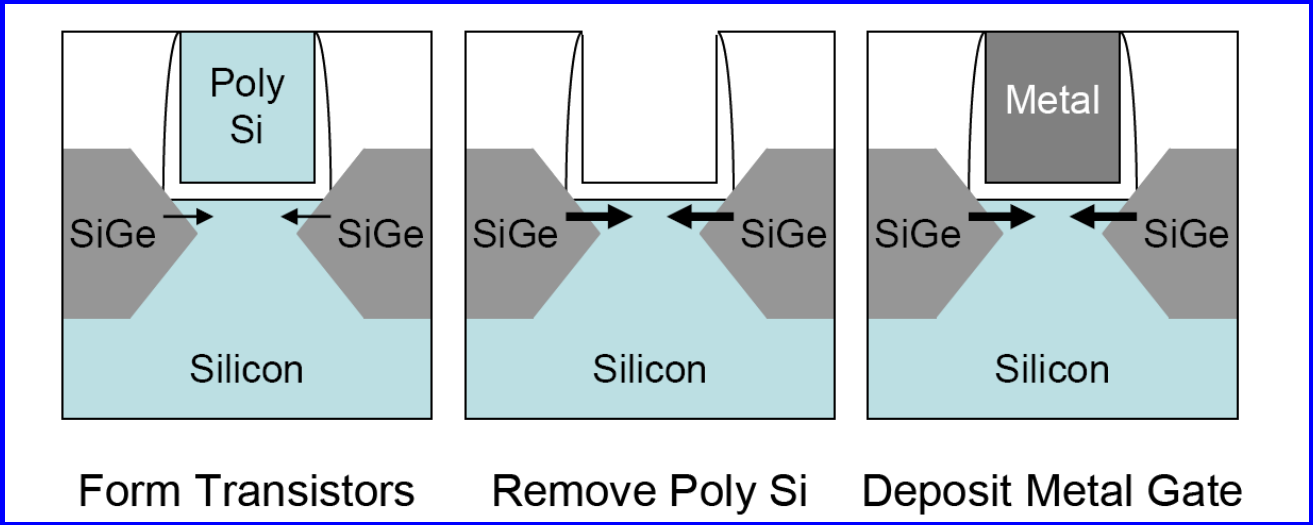


Thick top metal for power delivery and I/O routing

Key Device Features

- **30nm gate length with 112.5nm contacted gate pitch**
- **2nd generation high-k metal gate**
  - 0.9nm EOT
  - Replacement metal gate approach
    - Enables stress enhancement techniques
  - Replacement high-k approach
    - Improved performance
- **4th generation SiGe strained silicon PMOS device**
  - Increased Ge concentration
  - Closer proximity to channel for enhanced mobility
- **Raised NMOS S/D region**
  - Improved external resistance
- **2nd generation trench contacts**
  - Reduced contact resistance
  - Used as local interconnects

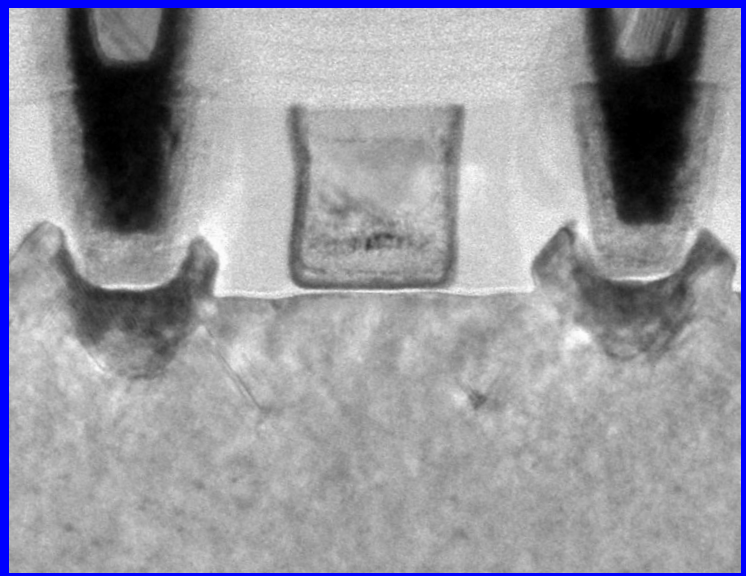
Replacement Metal Gate Benefits



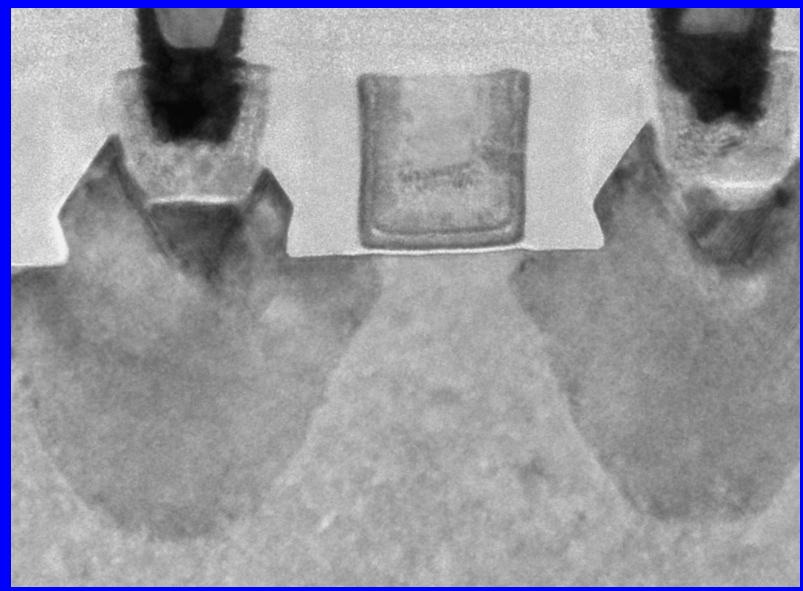
- High Thermal budget available for Midsection
  - Better Activation of S/D Implants
- Low thermal budget for Metal Gate
  - Large range of gate materials available
- **Significant enhancement of strain**
  - **Both NMOS and PMOS benefits**

Device Cross Sections

NMOS

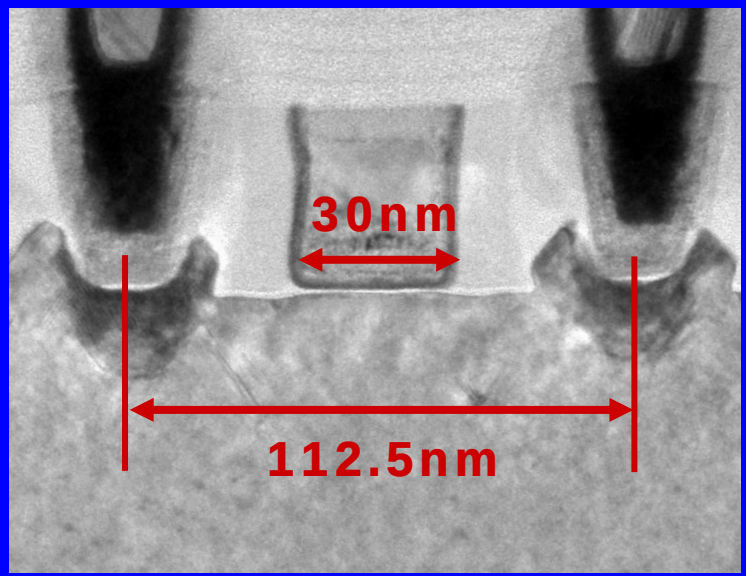


PMOS

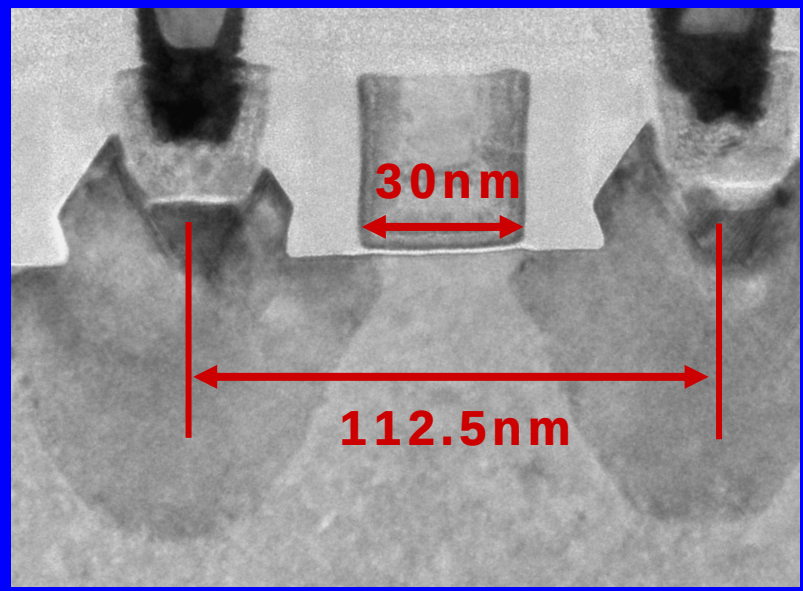


Device Cross Sections

NMOS



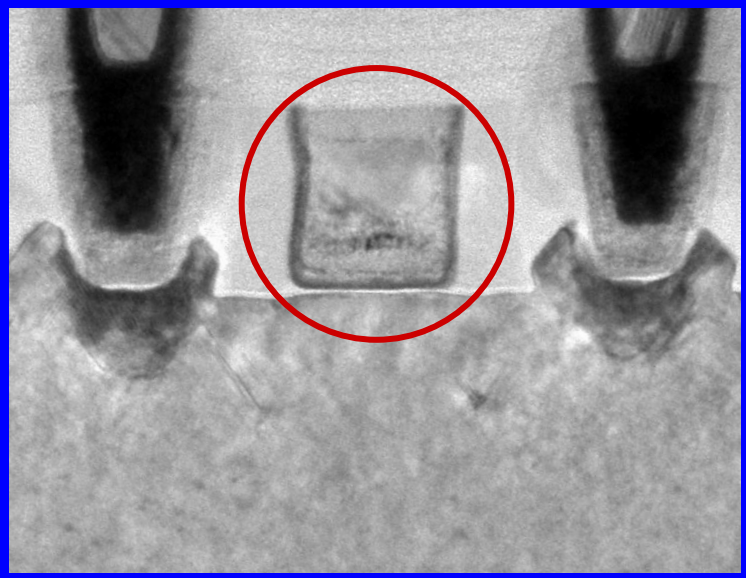
PMOS



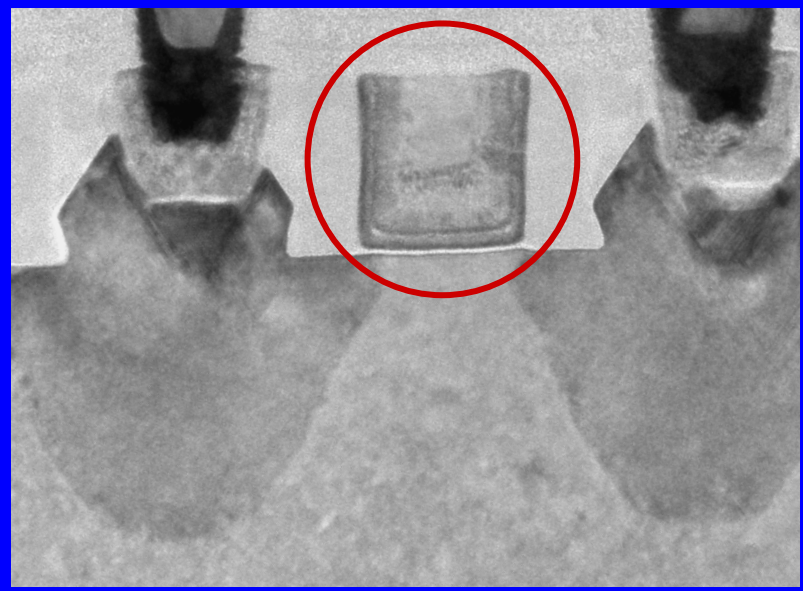
- 30nm gate length devices
- 112.5nm contacted gate pitch

Device Cross Sections

NMOS



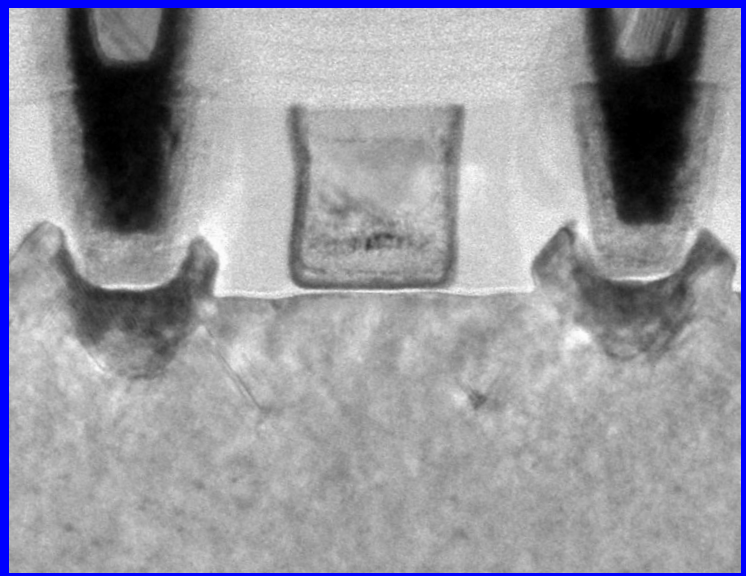
PMOS



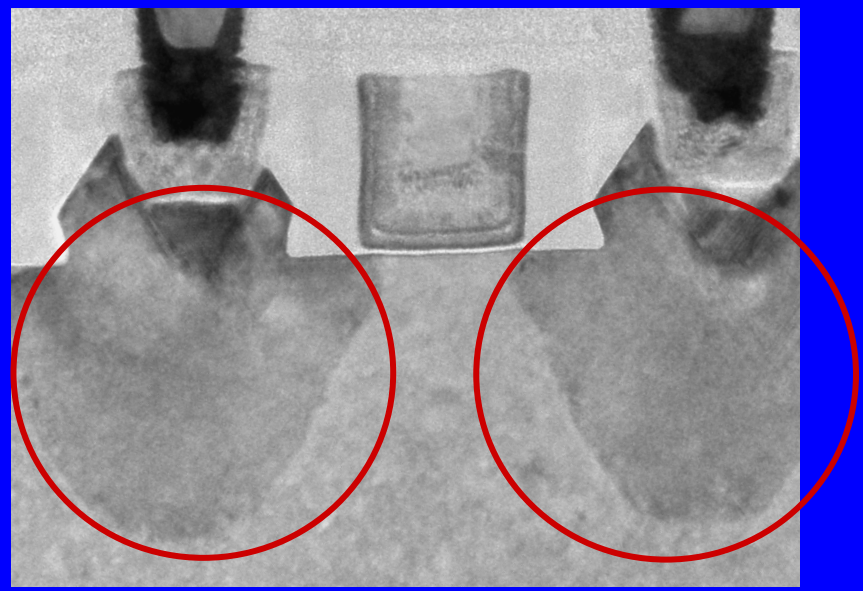
- 2nd generation high-k dual work function metal gate
- Replacement metal gate and high-k flow for improved performance

Device Cross Sections

NMOS



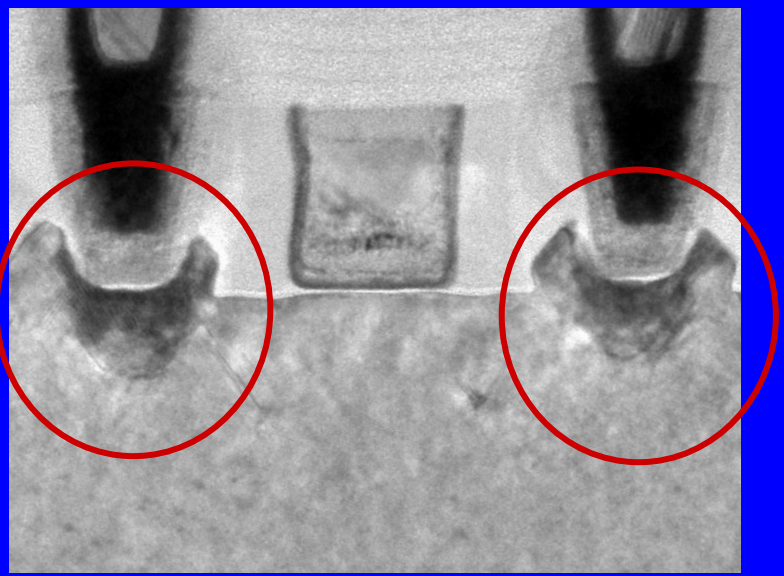
PMOS



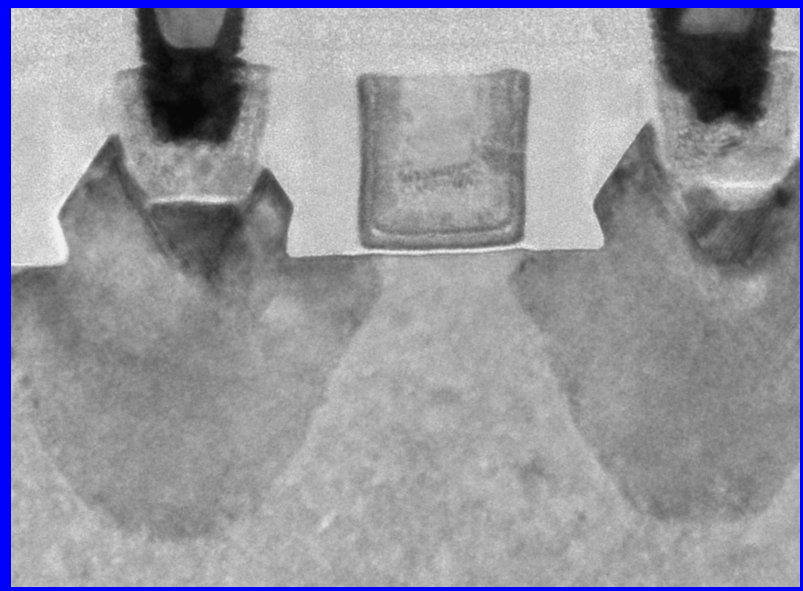
- 4th generation strained SiGe PMOS epi layer
- Increased Ge content and closer proximity to channel for higher strain

Device Cross Sections

NMOS



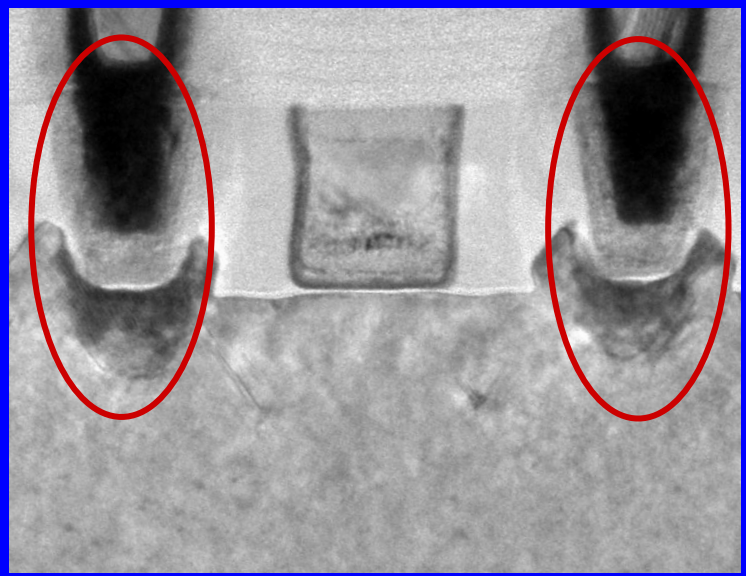
PMOS



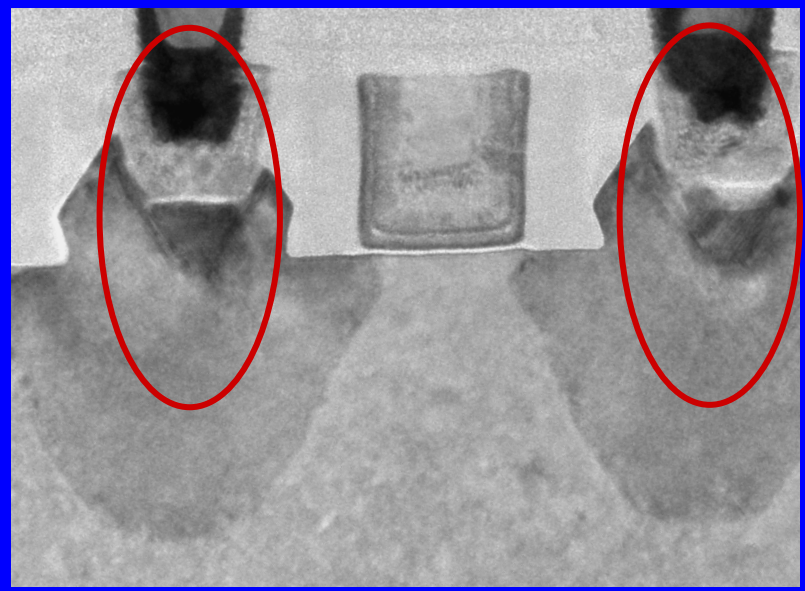
- Raised NMOS S/D region for improved Rext

Device Cross Sections

NMOS

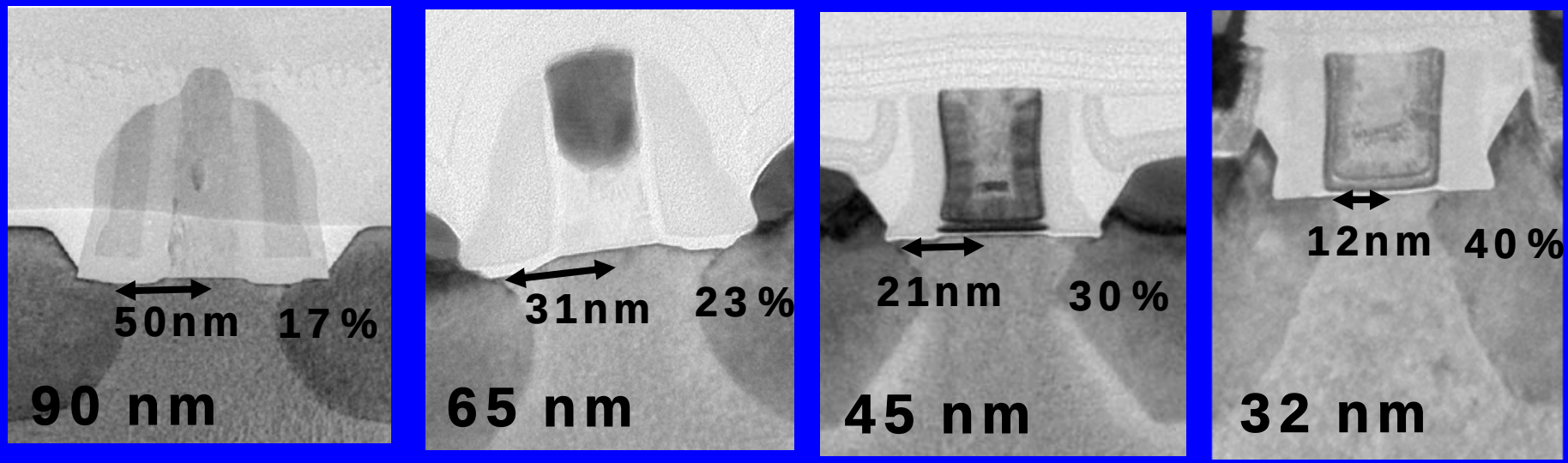


PMOS



- Dual trench contacts with tungsten lower layer and copper upper layer
- Trench contacts for improved resistance and local interconnects

PMOS Mobility Scaling



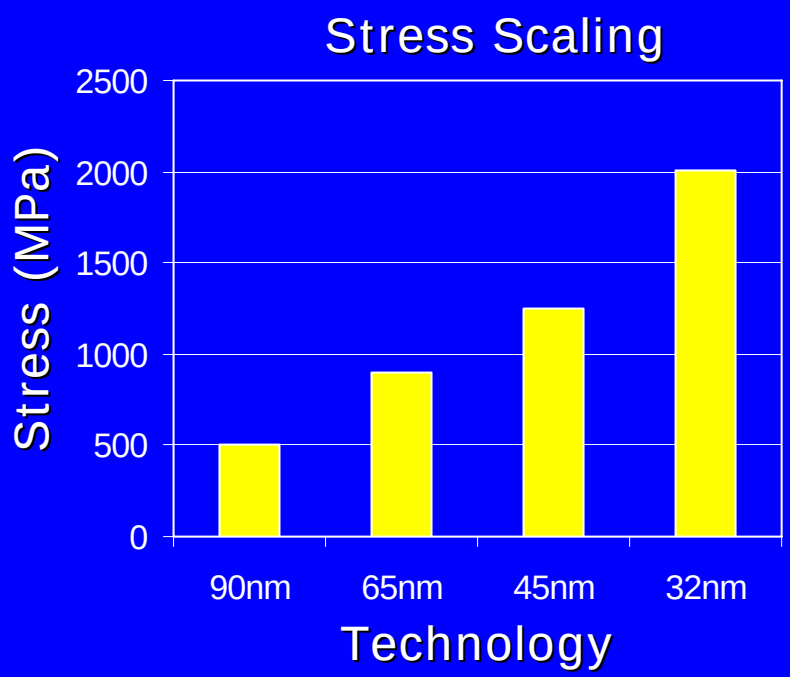
Increasing Ge Concentration

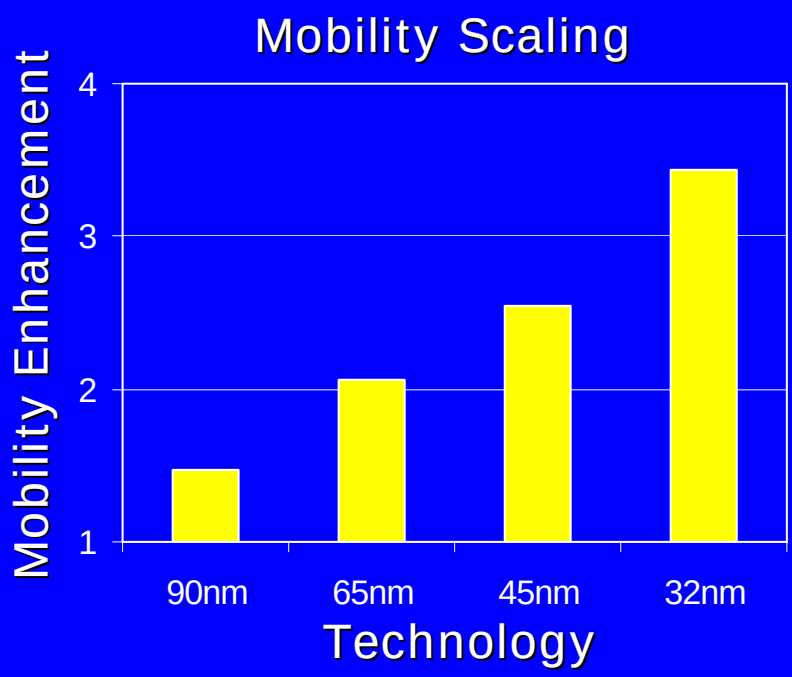
Closer Proximity

RMG Strain Enhancement

- Increasing Ge concentration and closer SiGe proximity improves drive
- Strain enhancement possible only in the RMG flow also improves drive

Stress and Mobility Enhancement



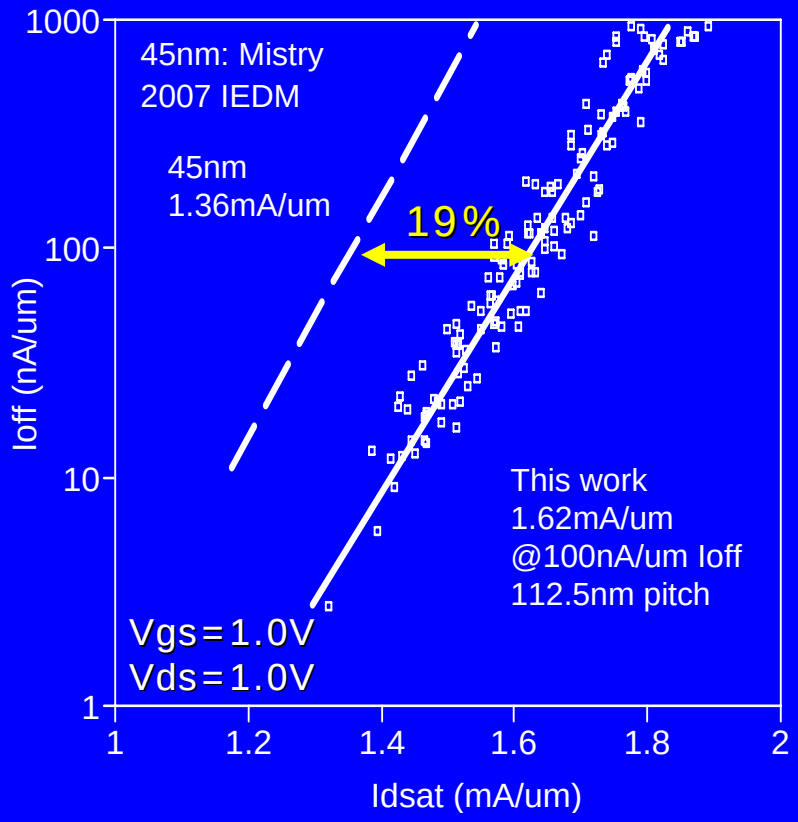


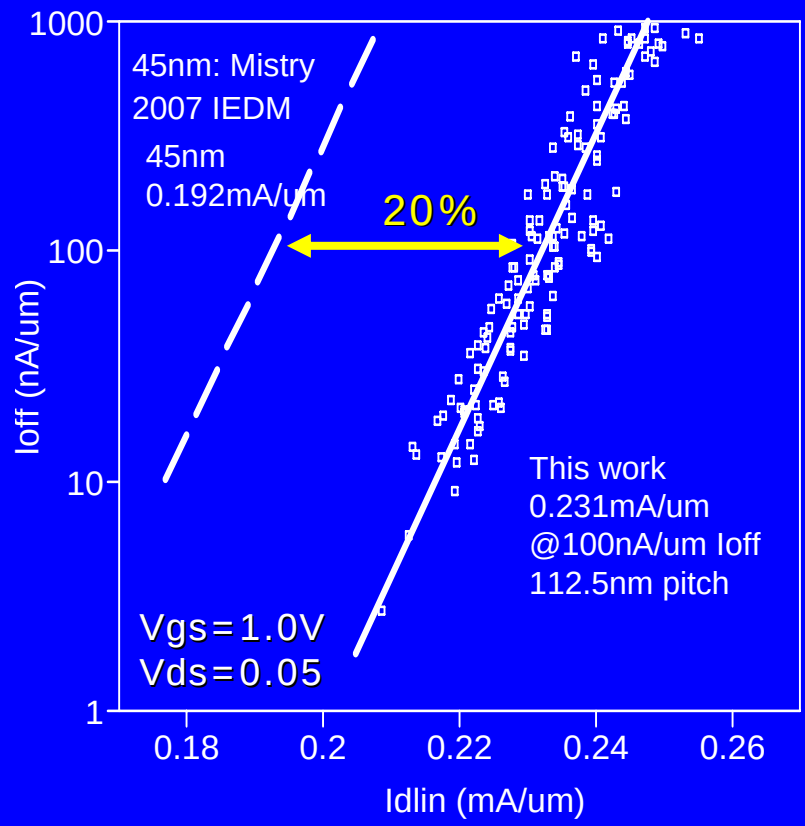
Innovations enable performance and pitch scaling

Outline

- Scaling Trends
- 32nm Technology Features
- **Device Performance**
- Variation and Vccmin
- Summary

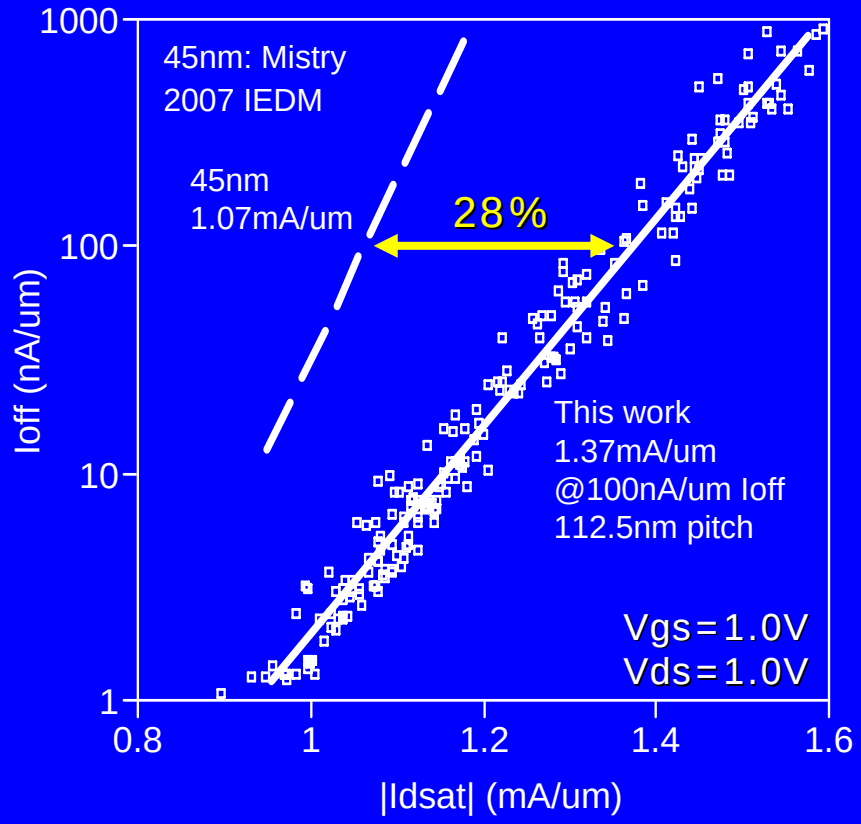
NMOS Performance

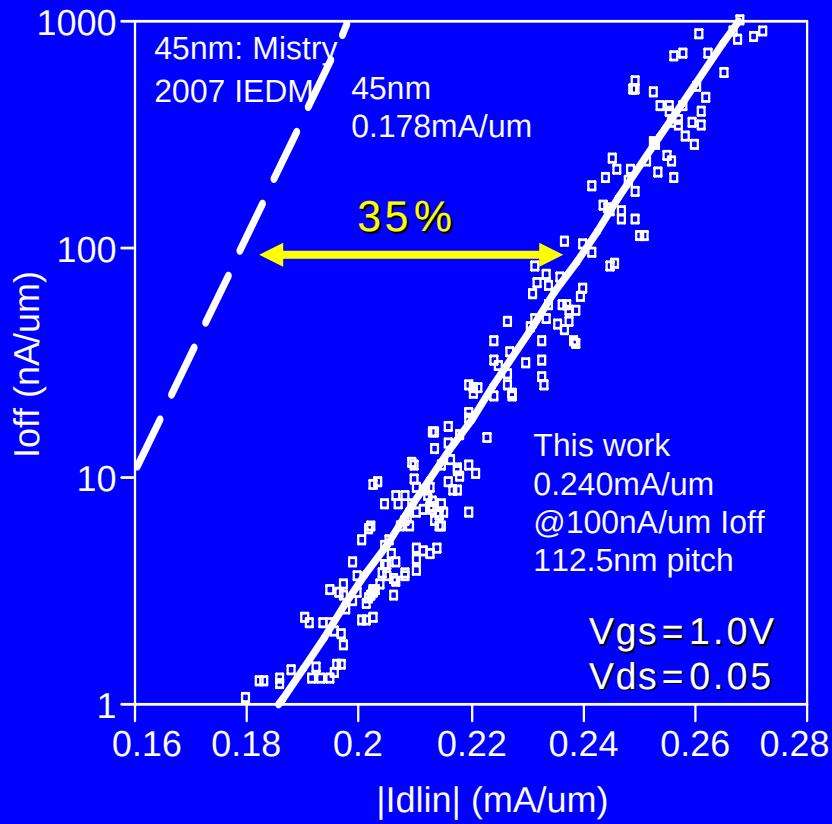




- 1.62mA/um Idsat and 0.231mA/um Idlin @ 100nA Ioff
  - 19% and 20% gain over 45nm

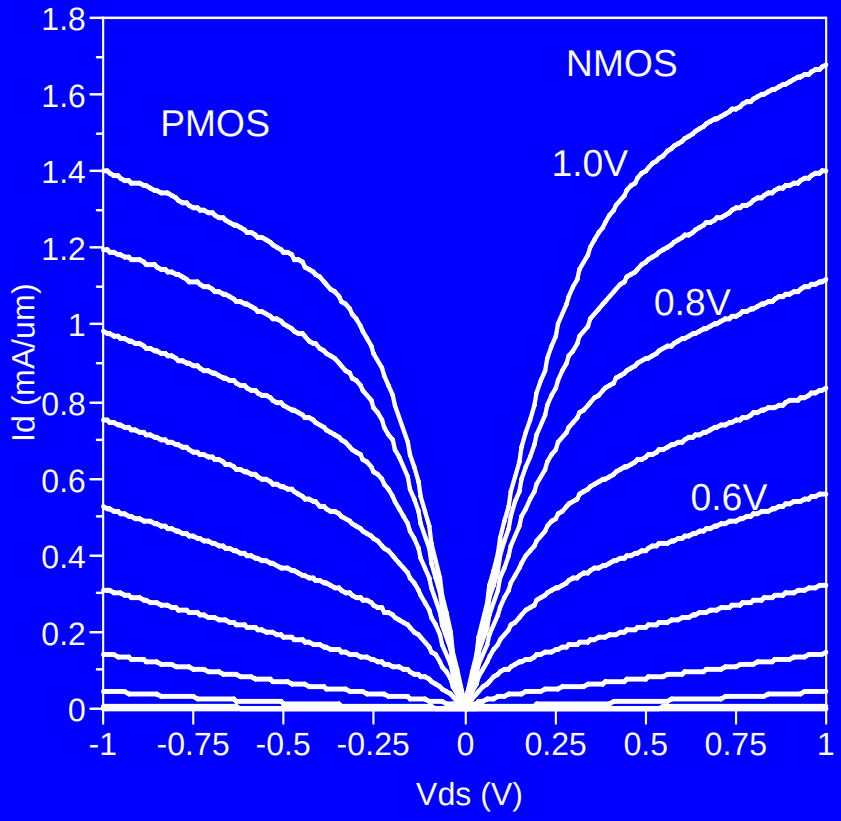
PMOS Performance

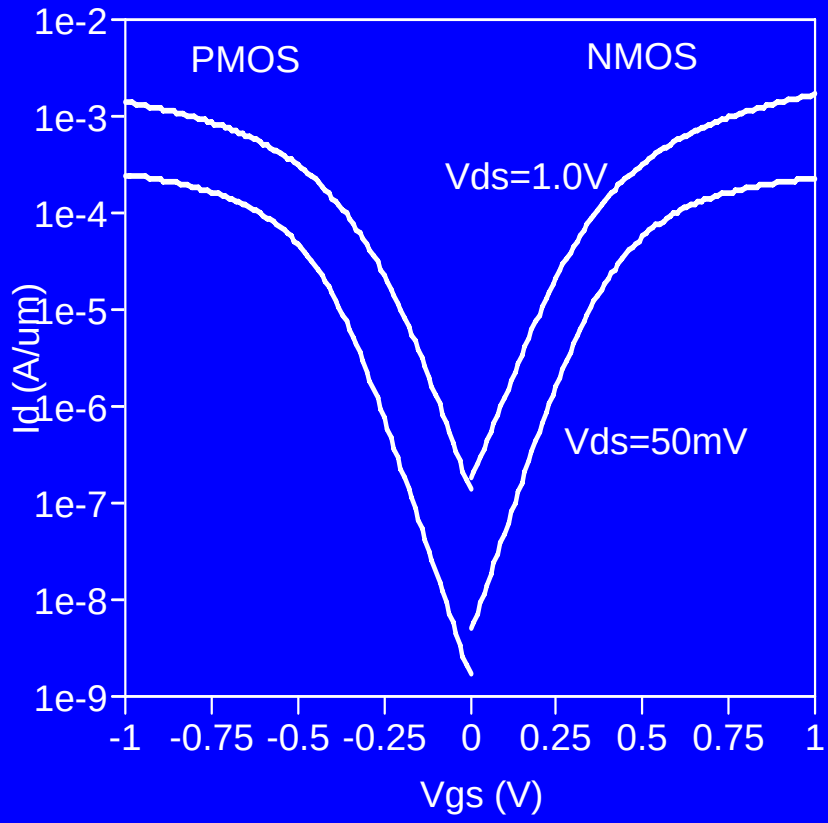




- 1.37mA/um Idsat and 0.240mA/um Idlin @ 100nA Ioff
- 28% and 35% gain over 45nm

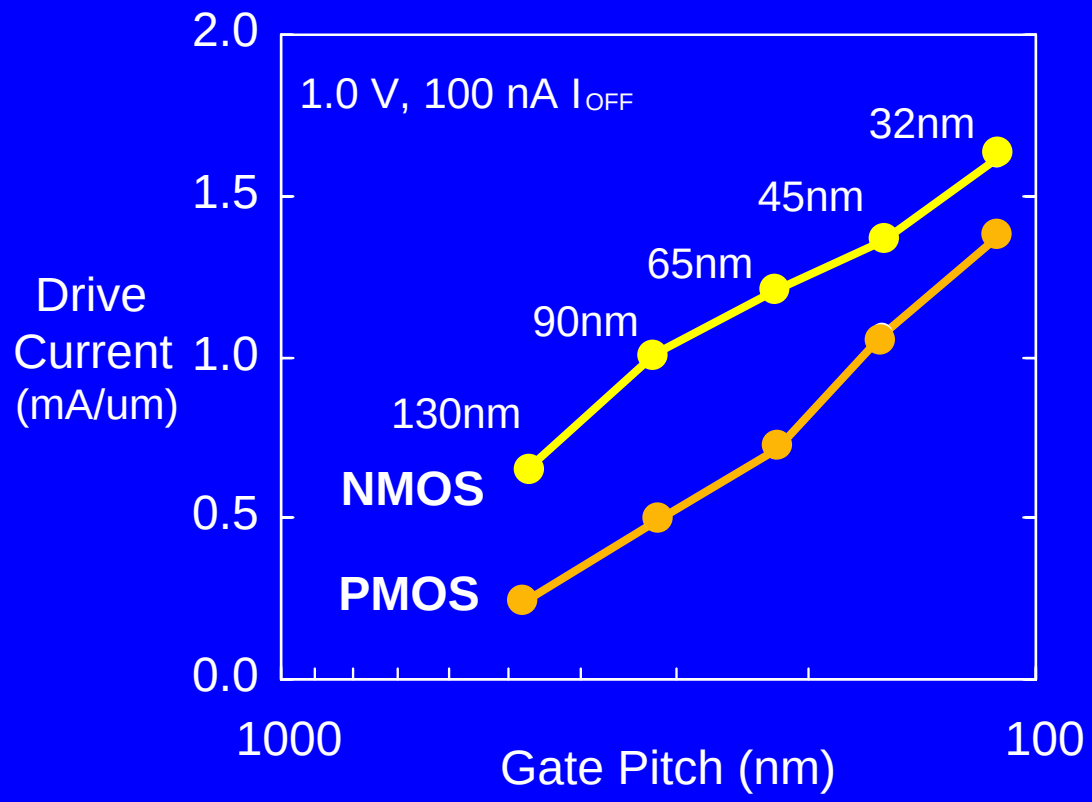
IV Characteristics





- PMOS drive currents are approaching NMOS values
- Subthreshold slopes are well maintained at ~100mV/decade

Pitch and Performance Scaling



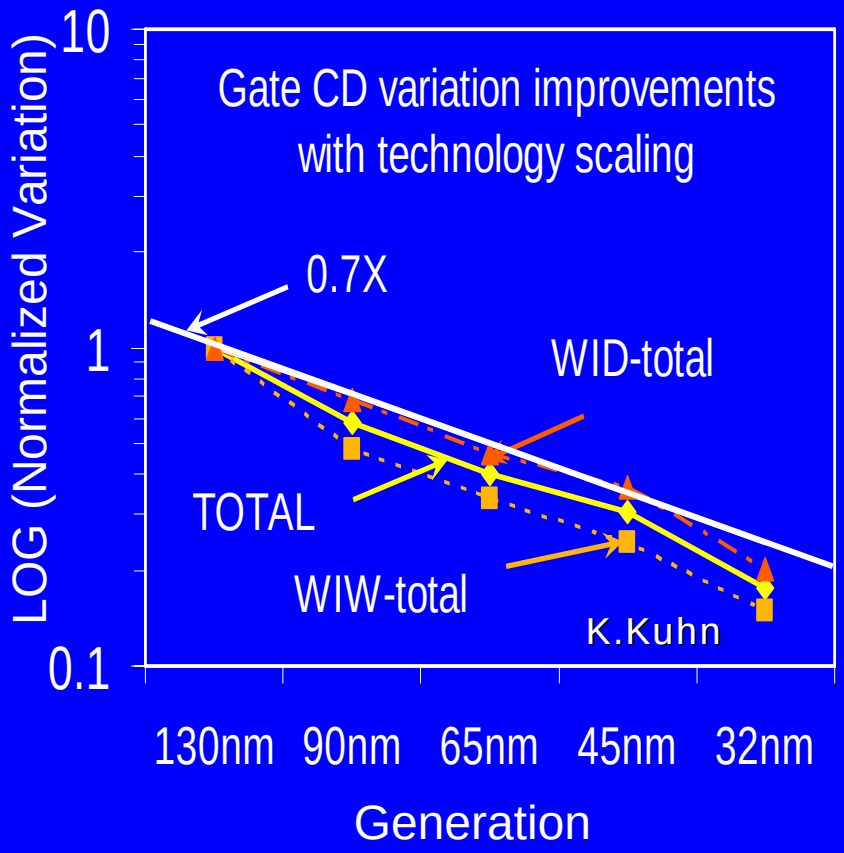
- Simultaneous performance and density improvement

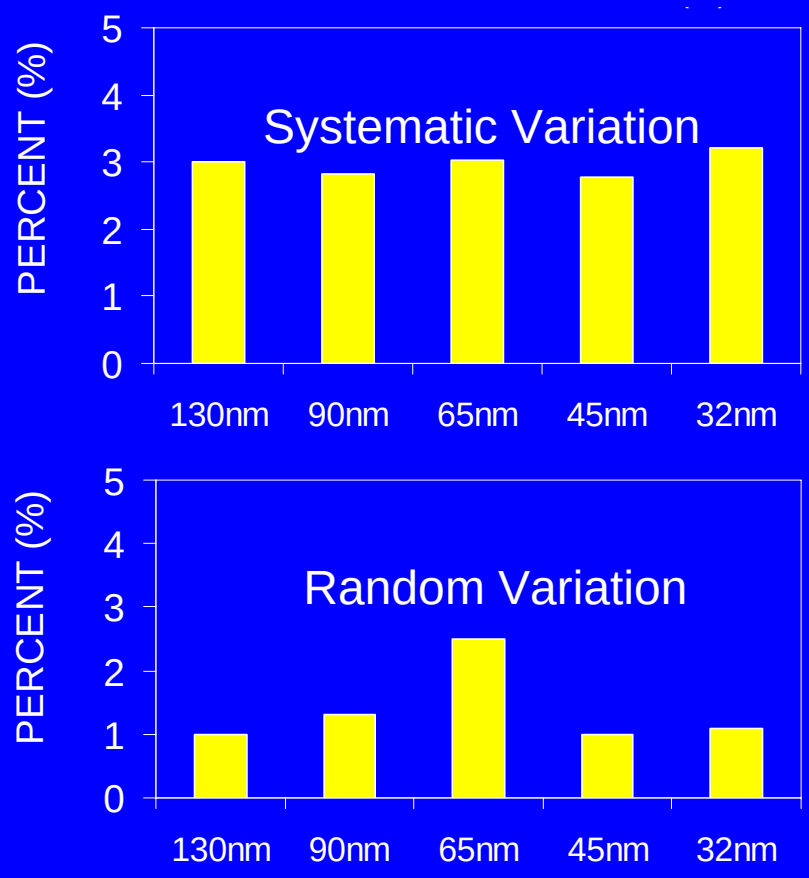
Highest reported drive currents at tightest reported gate pitch of any 32nm or 28nm technology

Outline

- Scaling Trends
- 32nm Technology Features
- Device Performance
- **Variation and Vccmin**
- Summary

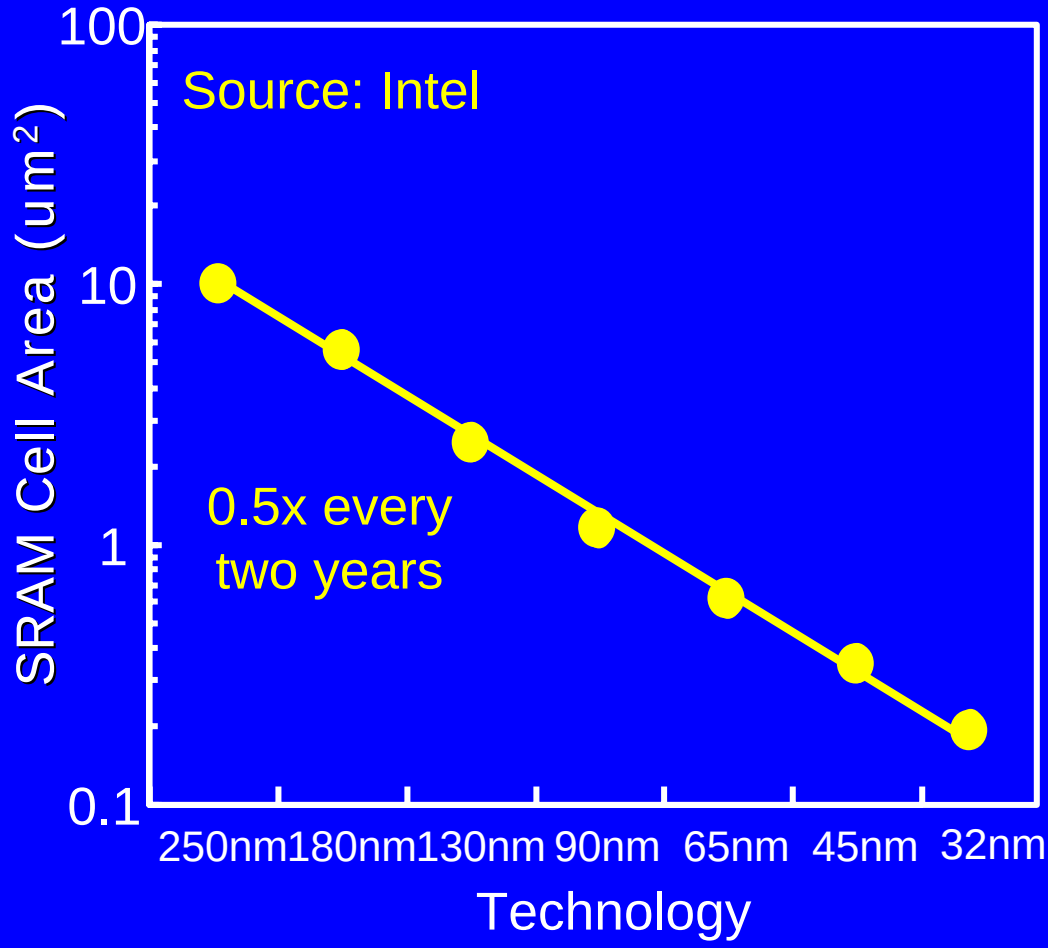
Device Variation

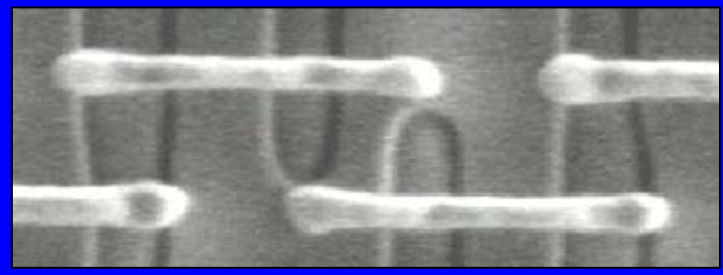




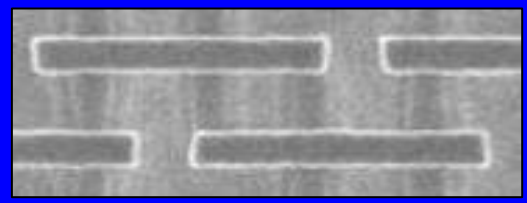
- 32nm technology continues the 0.7x reduction in CD variation
- No increase in systematic and random variation is seen

SRAM Scaling Trend

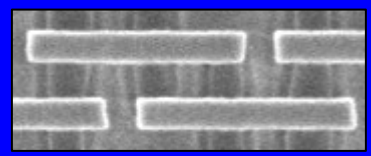




65 nm, 0.570 um²



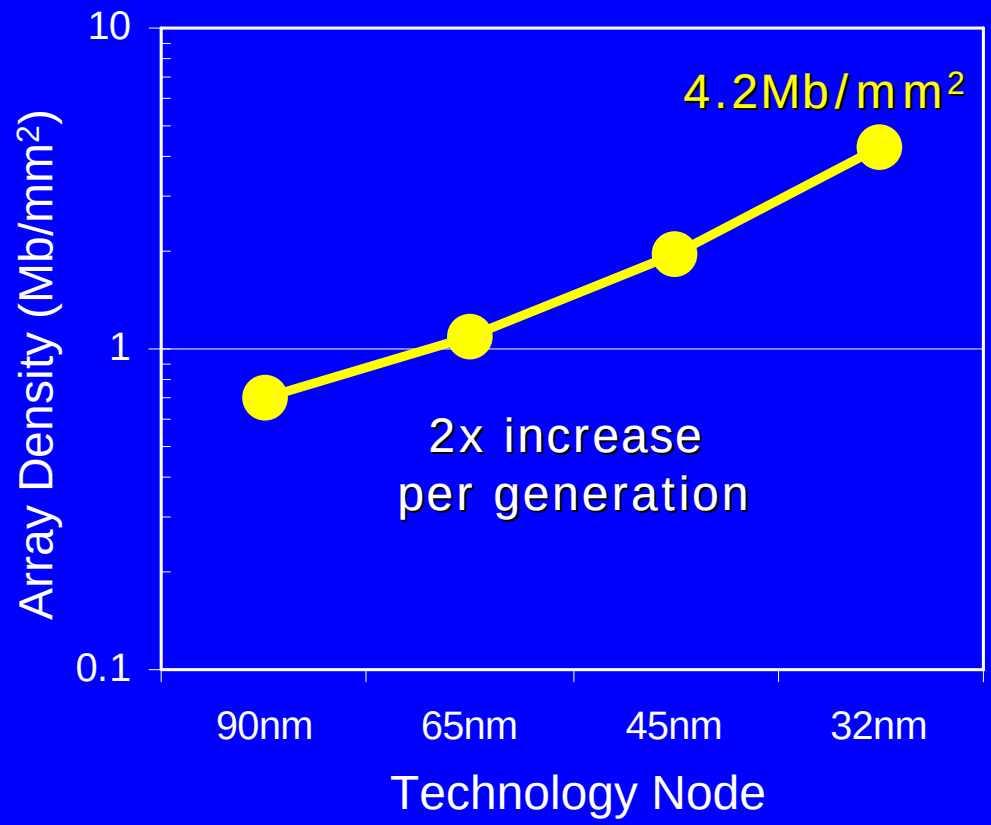
45 nm, 0.346 um²



32 nm, 0.171 um²

Transistor density doubles every two years

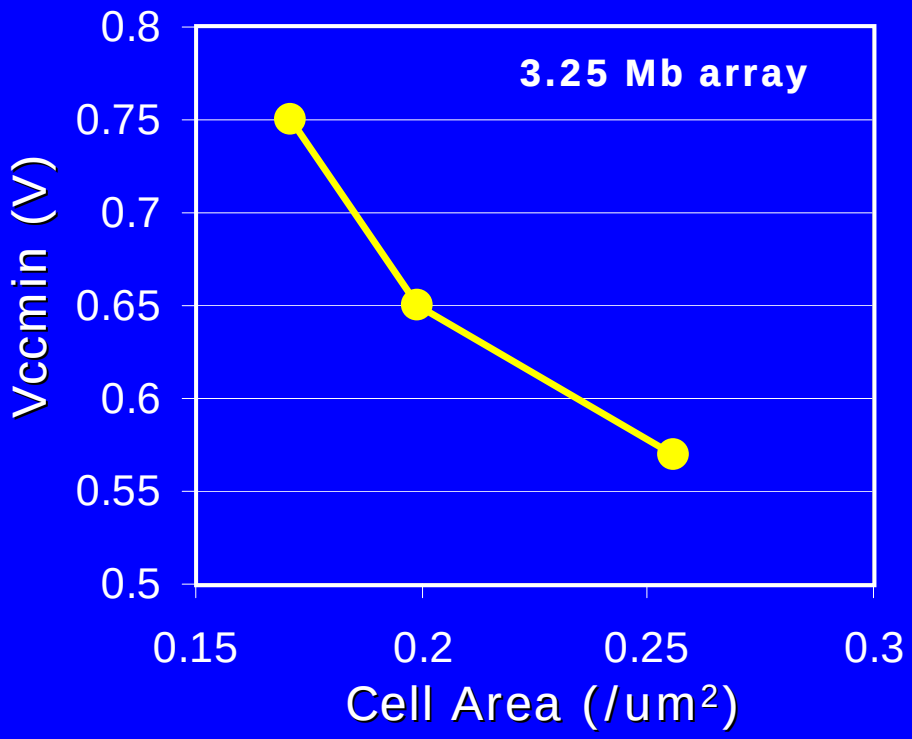
Array Density Scaling

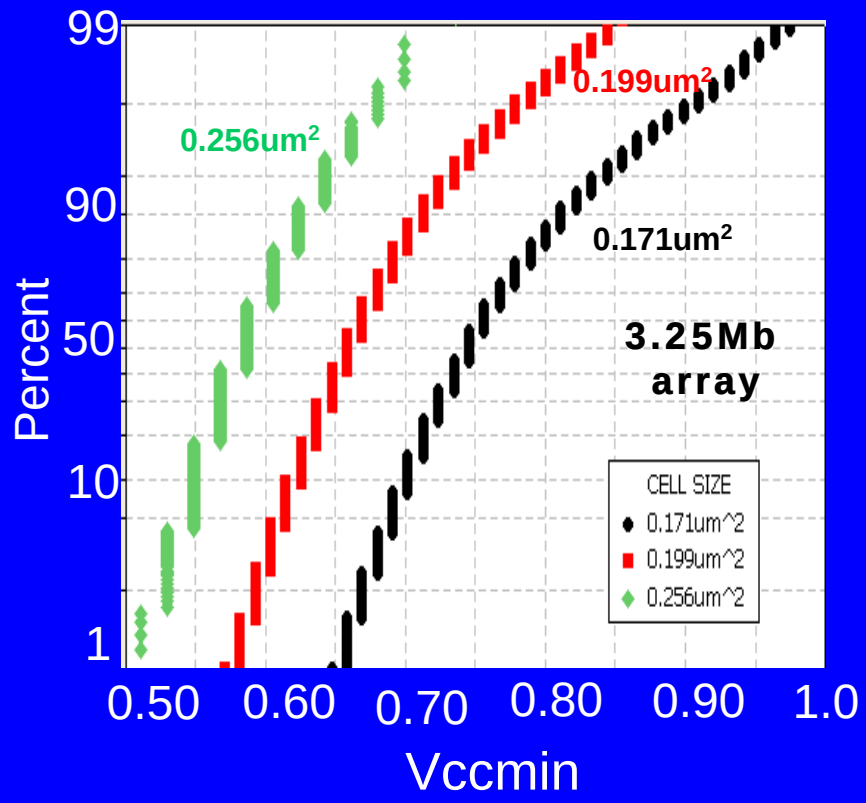


- Array density includes cells, sense amps and control circuitry

4.2Mbit/mm² array density is the highest reported for any 32nm or 28nm technology

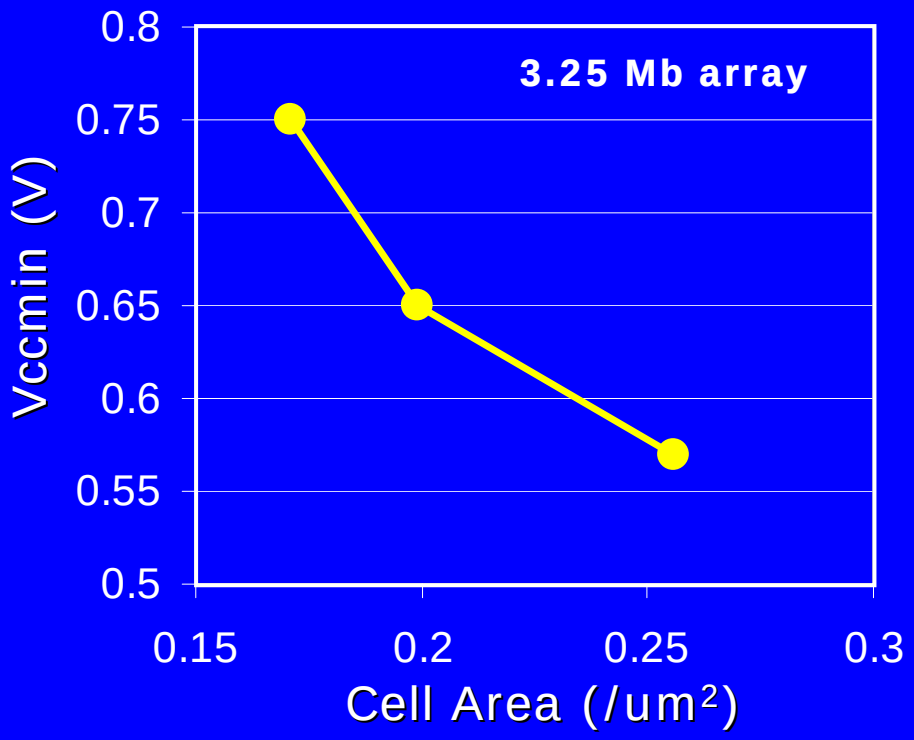
Cell Area and Vccmin

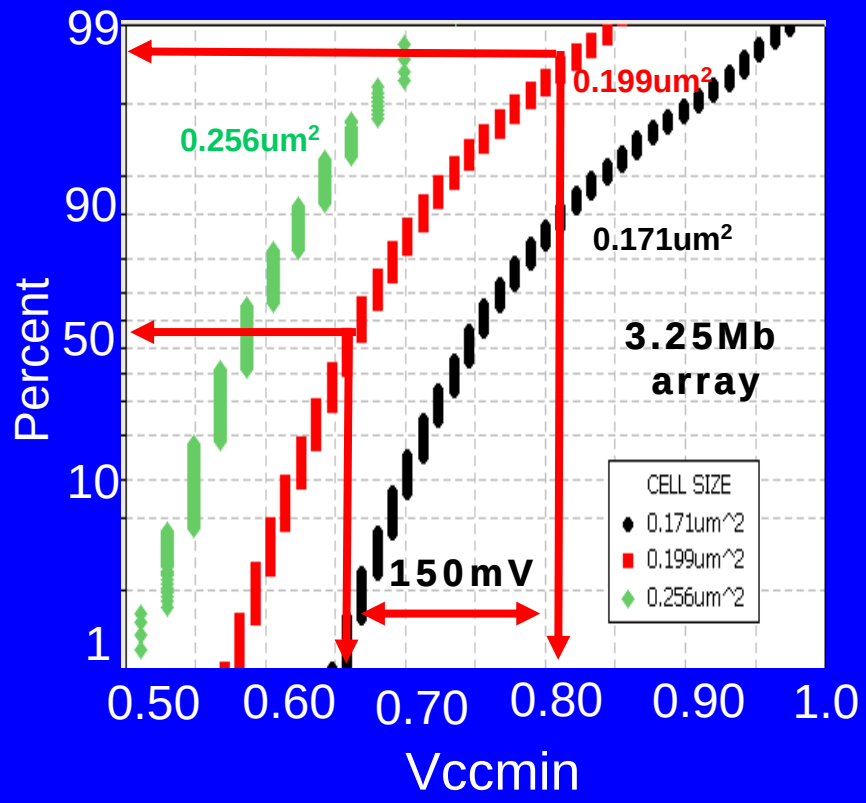




- As expected, larger SRAM cell sizes support lower Vccmin
- Vccmin depends on the distribution percentage reported

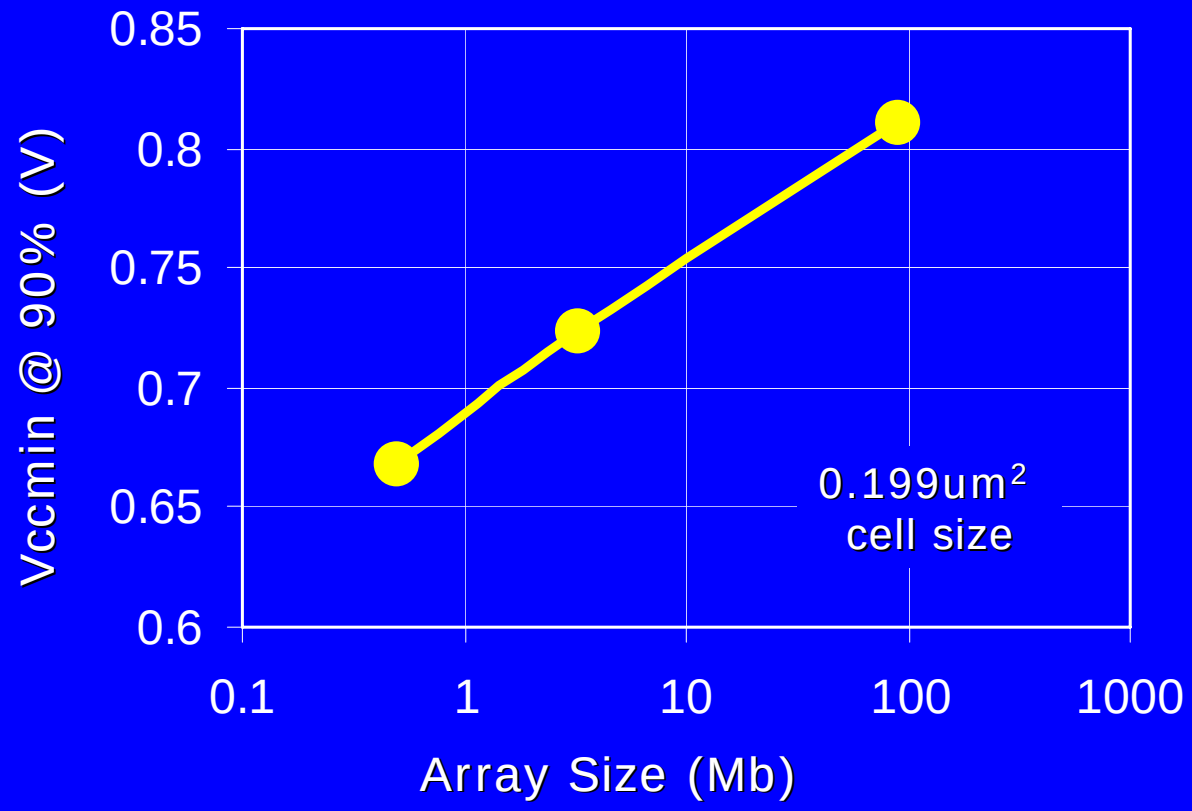
Cell Area and Vccmin





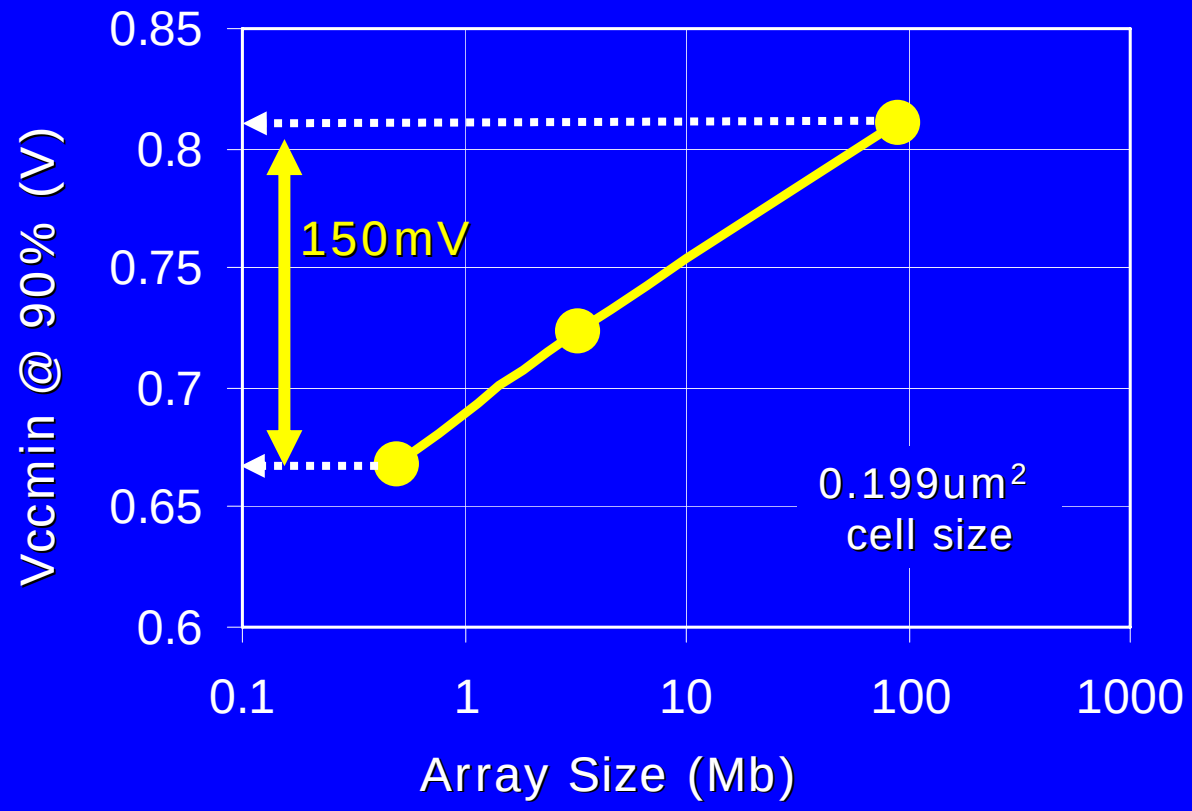
- As expected, larger SRAM cell sizes support lower Vccmin
- Vccmin depends on the distribution percentage reported

Cell Array Size and Vccmin



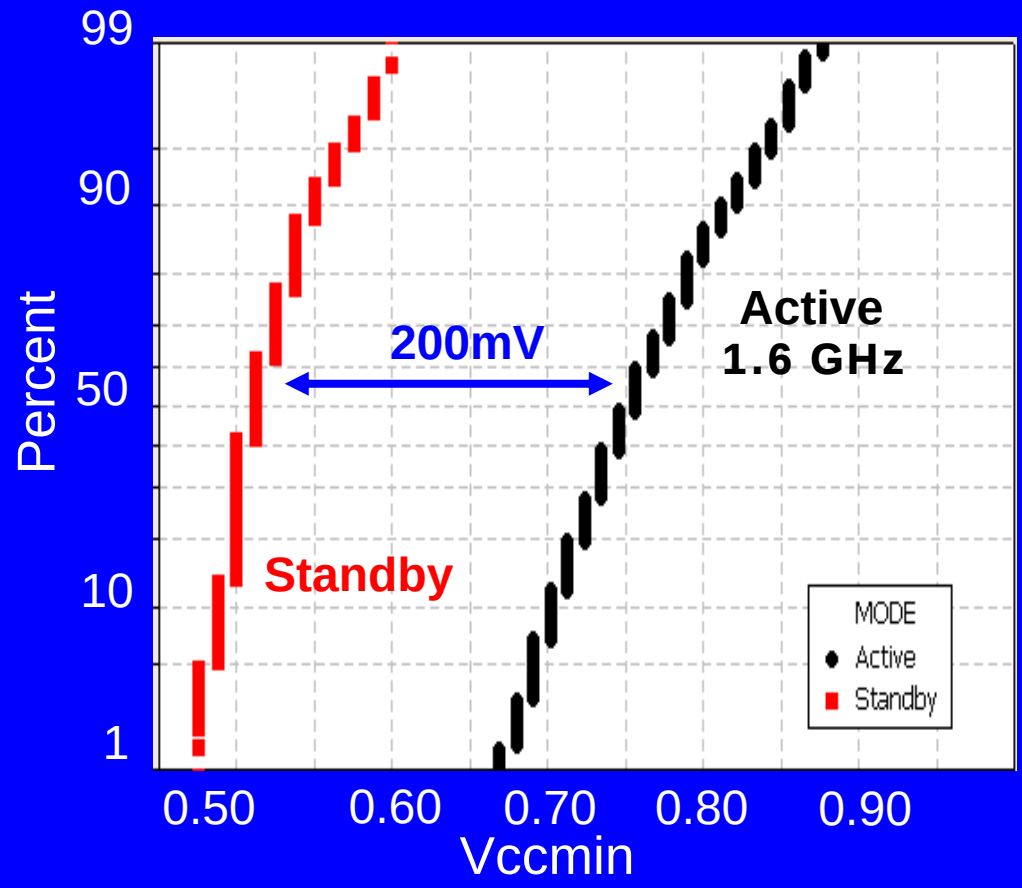
- Vccmin increases for larger array sizes due to the statistics from the larger number of cells

Cell Array Size and Vccmin



- Vccmin increases for larger array sizes due to the statistics from the larger number of cells

Active and Standby Vccmin

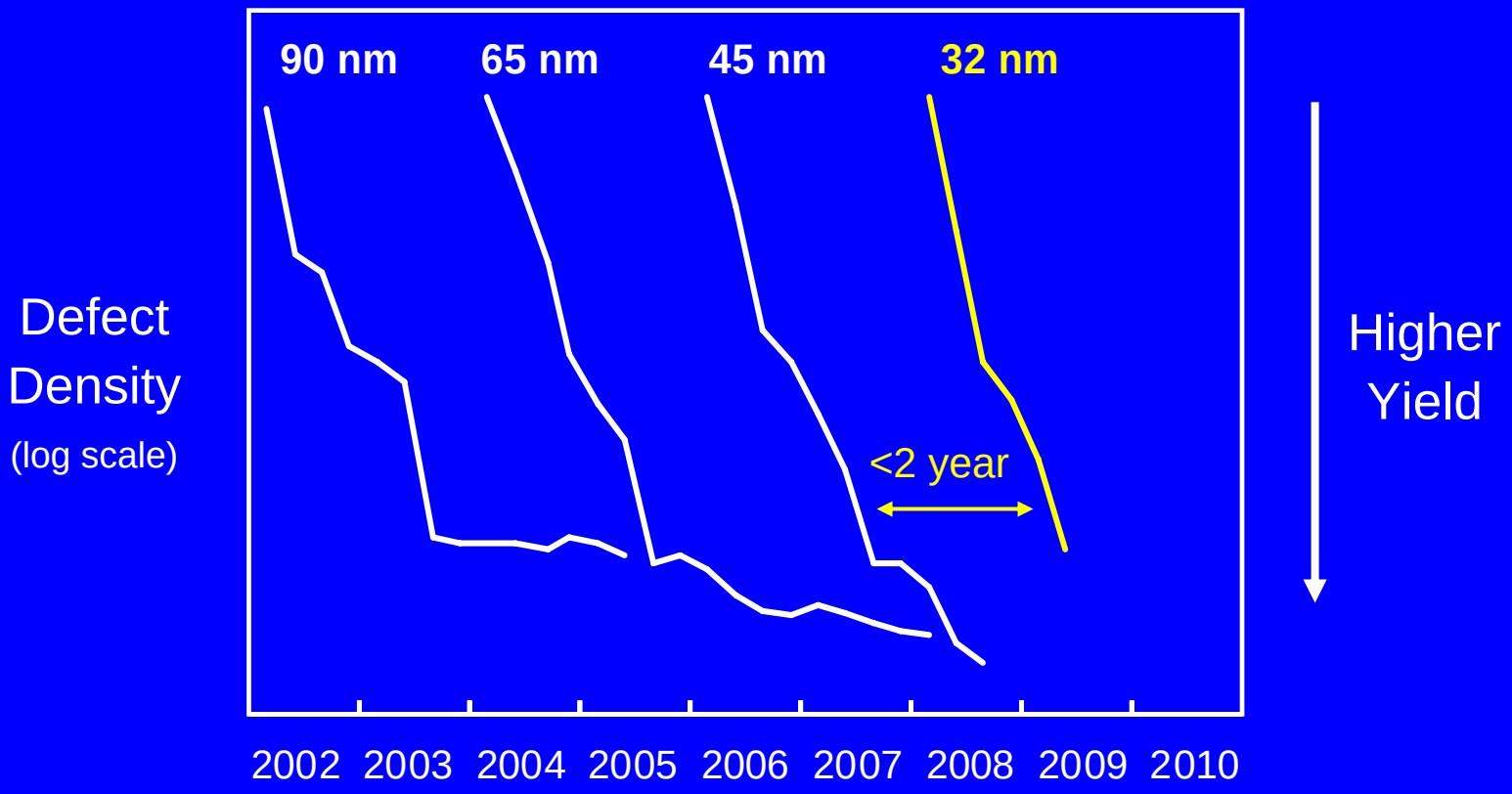


Cell size, array size, distribution and frequency must all be included when reporting Vccmin

Outline

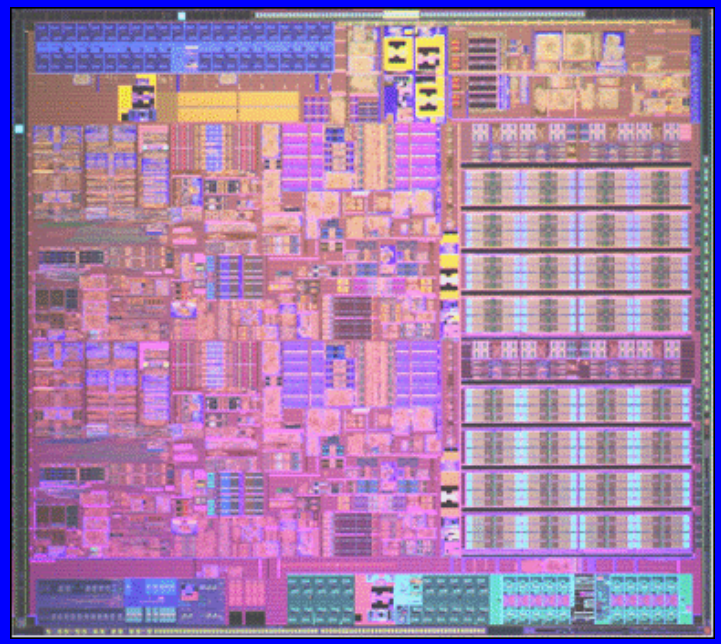
- Scaling Trends
- 32nm Technology Features
- Device Performance
- Variation and Vccmin
- **Summary**

32 nm Defect Density Trend



- Intel's 32 nm process has achieved the high yields needed for volume production
- 32 nm CPU products are presently being produced and shipped from two factories

32 nm Westmere Microprocessor



Dual core Westmere

Industry's first 32 nm processor

First in a family of 32 nm microprocessors based upon the Intel® microarchitecture codenamed Nehalem

Summary

- 32nm technology continues historic scaling trends
  - Reduced pitch and increased performance
- Record NMOS and PMOS drive currents
  - NMOS Idsat of 1.62mA/um @100nA/um Ioff, 1.0V
  - PMOS Idsat of 1.37mA/um @100nA/um Ioff, 1.0V
- Highest reported SRAM array density and tightest reported gate pitch for any 32nm or 28nm technology
- This 32nm technology is in high volume manufacturing of multi-core CPU products in multiple fabs

Acknowledgements

- The authors gratefully acknowledge the many people in the following organizations at Intel who contributed to this work:
  - Logic Technology Development
  - Quality and Reliability Engineering
  - Technology CAD
  - Components Research
  - Assembly & Test Technology Development

For further information on Intel's silicon technology, please visit our Technology & Research page at www.intel.com/technology